

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2005, 06:36:53 ; Search time 1663 Seconds
(without alignments)
874.117 Million cell updates/sec

Title: US-09-806-197-24

Perfect score: 30

Sequence: 1 aacaatgtatgcgggtacatcatgac 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_on.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_ats.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	30	100.0	2048	6	E40089 Plant promo
2	30	100.0	2048	6	E40090 Plant promo
3	30	100.0	2052	6	E40087 Plant promo
4	30	100.0	2052	6	E40093 Plant promo
5	30	100.0	2056	6	E40091 Plant promo
6	23.6	78.7	2042	6	AR076817 Sequence
7	23.6	78.7	2042	6	E15125 Promoter. 7
8	21.8	72.7	348283	2	AL125706 Rattus no
9	21.2	70.7	161762	2	CR352244 Danio rer
10	21.2	70.7	162325	2	CR774195 Danio rer
11	21.2	70.7	213033	2	AC110515 Mus muscu
12	21	70.0	140142	8	AC146937 Oryza sat
13	21	70.0	185386	10	AC025586 Genomic s
14	21	70.0	205668	10	AL592422 Mouse DNA
15	20.6	68.7	69674	9	AL136130 Human DNA
16	20.6	68.7	78026	2	AC023516 Homo sapi
17	20.6	68.7	215896	2	AC046180 Homo sapi
18	20.4	68.0	118001	9	AC073081 Homo sapi
19	20.4	68.0	193910	2	AC040996 Homo sapi

AC109822	Homo sapi
AC092430	Homo sapi
AC008034	Homo sapi
AC068315	Homo sapi
AC066599	Homo sapi
AC021996	Homo sapi
AC022382	Homo sapi
AC011610	Homo sapi
AC018829	Homo sapi
BX649535	Danio rer
AC125770	Rattus no
AC136192	Rattus no
AC023821	Homo sapi
AP004091	Oryza sat
AC139637	Mus muscu
AL606622	Oryza sat
AP004865	Oryza sat
AC142474	Mus muscu
AC151188	Bos tauru
AC128748	Rattus no
AC068650	Mus muscu
AC132648	Rattus no
AC073780	Mus muscu
AC128555	Rattus no
AC127884	Rattus no
AC133699	Rattus no
AC108524	Rattus no
AC109390	Rattus no
C0804634	Sequence
AC027665	Genomic s
AC095866	Rattus no
AL928993	Mouse DNA
AC069251	Genomic s
AL365199	Human DNA
AC119233	Mus muscu
AP001188	Homo sapi
AC092706	Homo sapi
AL591169	Homo sapi
AC117760	Mus muscu
AC139299	Mus muscu
AC101116	Rattus no
AC107804	Mus muscu
BX510341	Mus muscu
BX510341	Mus muscu
BX511227	Zebrafish
AC149978	Homo sapi
AC138382	Homo sapi
AC023101	Homo sapi
AC117902	Rattus no
AC093296	Homo sapi
AC022104	Homo sapi
AC144859	Mus muscu
AL772158	Zebrafish
AC150825	Callithr
AC027222	Homo sapi
AC103623	Mus muscu
AC103361	Mus muscu
AC114585	Mus muscu
AC090290	Mus muscu
AC136065	Rattus no
AC114410	Mus muscu
AC118815	Rattus no
AC027374	Homo sapi
BX649256	Danio rer
AC134450	Mus muscu
AC119913	Mus muscu
AC124532	Mus muscu
AC126059	Rattus no

[illegible]

```
Db      1742 AACAAATGATGTCGGGTGTACATCTATGAC 1771
RESULT 3
E40087
LOCUS      E40087      2052 bp      DNA      linear      PAT 31-JAN-2002
DEFINITION Plant promoter and terminator.
ACCESSION  E40087
VERSION    E40087.1 GI:18627203
KEYWORDS  JP 2000166577-A/1.
SOURCE     unidentified
ORGANISM   unclassified.
REFERENCE  1 (bases 1 to 2052)
AUTHORS   Nishikawa,S. and Oeda,K.
TITLE     Plant promoter and terminator
JOURNAL   Patent: JP 2000166577-A 1 20-JUN-2000;
          SUMITOMO CHEM CO LTD
COMMENT   OS Daucus carota L.
          PN JP 2000166577-A/1
          PD 20-JUN-2000
          PF 01-OCT-1999 JP 1999281475
          PR
          PI SATOMI NISHIKAWA, KENJI OEDA
          PC C12N15/09,A01H5/00,C12N1/21,C12N5/10/(C12N5/10,C12R1:91), PC
          C12N15/00,
          CC C12N5/00,(C12N5/00,C12R1:91)
          FH Key
          FT promoter Location/Qualifiers
             1..2052
             /organism="unidentified"
             /mol_type="genomic DNA"
             /db_xref="taxon:32644"

FEATURES             source
ORIGIN
Query Match      100.0%; Score 30; DB 6; Length 2052;
Best Local Similarity 100.0%; Pred. No. 0.00049;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AACAAATGATGTCGGGTGTACATCTATGAC 30
      ||||||||||||||||||||||||||||||||
Db      1746 AACAAATGATGTCGGGTGTACATCTATGAC 1775

RESULT 5
E40091
LOCUS      E40091      2056 bp      DNA      linear      PAT 31-JAN-2002
DEFINITION Plant promoter and terminator.
ACCESSION  E40091
VERSION    E40091.1 GI:18627207
KEYWORDS  JP 2000166577-A/5.
SOURCE     unidentified
ORGANISM   unclassified.
REFERENCE  1 (bases 1 to 2056)
AUTHORS   Nishikawa,S. and Oeda,K.
TITLE     Plant promoter and terminator
JOURNAL   Patent: JP 2000166577-A 5 20-JUN-2000;
          SUMITOMO CHEM CO LTD
COMMENT   OS Daucus carota L.
          PN JP 2000166577-A/5
          PD 20-JUN-2000
          PF 01-OCT-1999 JP 1999281475
          PR
          PI SATOMI NISHIKAWA, KENJI OEDA
          PC C12N15/09,A01H5/00,C12N1/21,C12N5/10/(C12N5/10,C12R1:91), PC
          C12N15/00,
          CC C12N5/00,(C12N5/00,C12R1:91)
          FH Key
          FT promoter Location/Qualifiers
             1..2056
             /organism="unidentified"
             /mol_type="genomic DNA"
             /db_xref="taxon:32644"

FEATURES             source
ORIGIN
Query Match      100.0%; Score 30; DB 6; Length 2056;
Best Local Similarity 100.0%; Pred. No. 0.00049;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 AACAAATGATGTCGGGTGTACATCTATGAC 30
      ||||||||||||||||||||||||||||||||
Db      1746 AACAAATGATGTCGGGTGTACATCTATGAC 1775

RESULT 6
AR076817
LOCUS      AR076817      2042 bp      DNA      linear      PAT 30-AUG-2000
DEFINITION Sequence 2 from patent US 5959176.
ACCESSION  AR076817
VERSION    AR076817
KEYWORDS  AR076817.1 GI:10003563
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 2042)
AUTHORS   Torikai,S. and Oeda,K.
TITLE     Plant promoter and utilization thereof
JOURNAL   Patent: US 5959176-A 2 28-SEP-1999;
```

FEATURES		Location/Qualifiers		REFERENCE				
source		1..2042		AUTHORS				
ORIGIN		/organism="unknown"						
		/mol_type="unassigned DNA"						
Query Match		78.7%; Score 23.6; DB 6; Length 2042;						
Best Local Similarity		86.7%; Pred. No. 0.95;						
Matches	26;	Conservative	0;	Mismatches	4;			
				Indels	0;			
				Gaps	0;			
QY		1 AACAACTGATGTCGGGTGATCATCTATGAC 30						
Db		1737 AACAACTGTTGTCGGGTGATATTATGAC 1766						
RESULT 7								
E15125		2042 bp		DNA				
LOCUS				linear				
DEFINITION				PAT 28-JUL-1999				
ACCESSION								
VERSION								
KEYWORDS								
SOURCE								
ORGANISM								
REFERENCE								
AUTHORS								
TITLE								
JOURNAL								
COMMENT								
OS		Daucus carota L. (carrot)						
PN		JP 1998052273-A/2						
PD		24-FEB-1998						
PF		12-AUG-1996 JP 1996212680						
PI		TORIKAI TOSHIMI, OITA KENJI						
PC		C12N15/09,A01H5/00,C07H21/04,C07K14/415,C12N1/21,C12N5/10; CC						
CC		strandedness: Single;						
CC		topology: Linear;						
CC		hypotheetical: No;						
FH		anti-sense: No;						
FT		Location/Qualifiers						
FT		1..2042						
FT		source						
FT		/organism="Daucus carota L."						
FT		promoter						
FT		1<..<2042.						
FEATURES		Location/Qualifiers						
source		1..2042						
		/organism="unidentified"						
		/mol_type="genomic DNA"						
		/db_xref="taxon:32644"						
ORIGIN								
Query Match		78.7%; Score 23.6; DB 6; Length 2042;						
Best Local Similarity		86.7%; Pred. No. 0.95;						
Matches	26;	Conservative	0;	Mismatches	4;			
				Indels	0;			
				Gaps	0;			
QY		1 AACAACTGATGTCGGGTGATCATCTATGAC 30						
Db		1737 AACAACTGTTGTCGGGTGATATTATGAC 1766						
RESULT 8								
AC125706		348283 bp		DNA				
LOCUS				linear				
DEFINITION				HTG 13-NOV-2002				
ACCESSION								
VERSION								
KEYWORDS								
SOURCE								
ORGANISM								
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;								
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;								

Rattus.

1 (bases 1 to 348283)

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,B., Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,K., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Garcia,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,D., Kovar,C., Kowis,C., Kraft,C.F., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhera,L., Loulseghe,H., Lozano,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Mirja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,N., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokeme,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pul.L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Speed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willison,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 348283)

Worley,K.C.

Direct Submission

Submitted (29-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 348283)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 13, 2002 this sequence version replaced gi:23832296.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequencing contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GDZZ

Center clone name: CH230-11J9

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 272438 bases at least Q40

Consensus quality: 277533 bases at least Q30

Consensus quality: 280756 bases at least Q20

Estimated insert size: 302017; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 8 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will

* be preserved.

* 1 11765: contig of 11765 bp in length

* 11766: gap of unknown length

* 11866: contig of 37002 bp in length

* 48867: gap of unknown length

* 48967: gap of unknown length

* 54443: contig of 5476 bp in length

* 54543: gap of unknown length

* 54544: contig of 5059 bp in length

* 59603: gap of unknown length

* 343138: contig of 283436 bp in length

* 343238: gap of unknown length

* 344661: contig of 1423 bp in length

* 344662: gap of unknown length

* 345971: contig of 1210 bp in length

* 345972: gap of unknown length

* 346072: contig of 2212 bp in length.

FEATURES

Source

1. .348283

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-11J9"

1. .1925

/note="wgs_end_extension"

clone_end:Sp6

1909. .2527

/note="clone boundary"

clone_end:Sp6

end sequence:BH342645"

8272. .11765

/note="wgs contig"

11866. .15119

/note="wgs contig"

20801. .22623

/note="wgs contig"

22674. .27148

/note="wgs contig"

48968. .52962

/note="wgs contig"

53013. .54443

/note="wgs contig"

misc_feature 54544. .56816
/note="wgs contig"
misc_feature 56870. .59602
/note="wgs contig"
misc_feature 59703. .61990
/note="wgs contig"
misc_feature 68117. .71973
/note="wgs contig"
misc_feature 338673. .340201
/note="wgs contig"
misc_feature 340622. .343138
/note="wgs contig"

ORIGIN

Query Match 72.7%; Score 21.8; DB 2; Length 348283;

Best Local Similarity 92.0%; Pred. No. 12;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 AACATGATATGTCGGGTGTACATCT 25

Db 90808 AACAAATGATATGTCAGGTGTACGTCT 90832

RESULT 9

CR352244/c

LOCUS

DEFINITION

CR352244 161762 bp DNA linear HTG 27-MAR-2004

unordered pieces.

CR352244

CR352244.4 GI:45772227

HTG; HTGS PHASE1.

SOURCE

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 161762)

Sims,S.

Direct Submission

Submitted (26-MAR-2004) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Mar 27, 2004 this sequence version replaced gi:45598726.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfish-help@sanger.ac.uk

----- Project Information

Center project name: zcl31e11

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 157149 bases at least Q40

Consensus quality: 157757 bases at least Q30

Consensus quality: 158375 bases at least Q20

Insert size: 160462; sum-of-contigs

Insert size: 174990; 4.8% error; agarose-fp

Quality coverage: 8.14x in Q20 bases; sum-of-contigs Quality

coverage: 7.67x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 14 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 16722: contig of 16722 bp in length

* 16723 16822: gap of 100 bp

* 16823 28393: contig of 11571 bp in length

* 28394 28493: gap of 100 bp

```

* 28494 40368: contig of 11875 bp in length
* 40369 40468: gap of 100 bp
* 40469 57841: contig of 17373 bp in length
* 57841 57941: gap of 100 bp
* 57941 69071: contig of 11130 bp in length
* 69071 69171: gap of 100 bp
* 69171 77218: contig of 8047 bp in length
* 77218 77318: gap of 100 bp
* 77318 85619: contig of 8301 bp in length
* 85619 85720: gap of 100 bp
* 85720 89373: contig of 3654 bp in length
* 89373 89473: gap of 100 bp
* 89473 98585: contig of 9112 bp in length
* 98585 98685: gap of 100 bp
* 98685 103814: contig of 5129 bp in length
* 103814 103915: gap of 100 bp
* 103915 110518: contig of 6603 bp in length
* 110518 110618: gap of 100 bp
* 110618 139606: contig of 28989 bp in length
* 139606 139707: gap of 100 bp
* 139707 144556: contig of 4850 bp in length
* 144556 144657: gap of 100 bp
* 144657 161762: contig of 17106 bp in length.
FEATURES
    source
        1..161762
            /organism="Danio rerio"
            /mol_type="genomic DNA"
            /db_xref="taxon:7955"
            /clone="CH211-131E11"
            /clone_lib="CHORI-211"
        1..16722
            /note="assembly fragment:01696
            fragment_chain:1
            clone_end:SP6
            vector_side:left"
        16823..28393
            /note="assembly fragment:00818
            fragment_chain:1"
        28494..40368
            /note="assembly fragment:00996
            fragment_chain:1"
        40469..57841
            /note="assembly fragment:01195
            fragment_chain:1"
        57942..69071
            /note="assembly fragment:00668
            fragment_chain:1"
        69172..77218
            /note="assembly fragment:00429
            fragment_chain:1"
        77319..85619
            /note="assembly fragment:00216
            fragment_chain:1"
        85720..89373
            /note="assembly fragment:00050
            fragment_chain:1"
        89474..98585
            /note="assembly fragment:00542
            fragment_chain:1"
        98686..103814
            /note="assembly fragment:00090
            fragment_chain:1"
        103915..110517
            /note="assembly fragment:00319
            fragment_chain:1"
        110618..139606
            /note="assembly fragment:02011
            fragment_chain:2"
        139707..144556
            /note="assembly fragment:00147
            fragment_chain:2"
        144657..161762
            /note="assembly fragment:01414

```

```

fragment_chain:2
clone_end:T7
vector_side:right"

ORIGIN
    Query Match          70.7%;   Score 21.2;   DB 2;   Length 161762;
    Best Local Similarity 88.5%;   Pred.No.23;
    Matches 23;   Conservative 0;   Mismatches 3;   Indels 0;   Gaps 0;

Qy 4 AATGTATGTCGGGTGACATCTATGA 29
    ||||| ||||| ||||| ||||| |||||
Db 9417 AATGTGTGTCGGGTGTCATCTGTGA 9392

RESULT 10
CR774195
LOCUS
DEFINITION
    CR774195 162325 bp DNA linear HTG 16-SEP-2004
    Danio rerio clone DKEY-73P2, *** SEQUENCING IN PROGRESS ***, 7
    unordered pieces.
ACCESSION
    CR774195
VERSION
    CR774195.1 GI:52213980
KEYWORDS
    HTG; HTGS_PHASE1.
SOURCE
    Danio rerio (zebrafish)
    ORGANISM
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
        Cypriniformes; Cyprinidae; Danio.
        McClay, K.
        Direct Submission
        Submitted (15-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
        Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
        zfsh-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
        ----- Genome Center
        Center: Wellcome Trust Sanger Institute
        Center code: SC
        Web site: http://www.sanger.ac.uk
        Contact: zfsh-help@sanger.ac.uk
        ----- Project Information
        Center project name: ZK73P2
        ----- Summary Statistics
        Assembly program: XGAP4; version 4.5
        Chemistry: Dye-terminator; 100% of reads
        Consensus quality: 160506 bases at least Q40
        Consensus quality: 160814 bases at least Q30
        Consensus quality: 161069 bases at least Q20
        Insert size: 161725; sum-of-contigs
        Insert size: 167136; 0.4% error; agarose-fp
        Quality coverage: 8.68x in Q20 bases; sum-of-contigs Quality
        coverage: 8.41x in Q20 bases; agarose-fp
        -----
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 7 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence.
        * as soon as it is available and the accession number will
        * be preserved.
        * 1 17481: contig of 17481 bp in length
        * 17482 17581: gap of 100 bp
        * 17582 38294: contig of 20713 bp in length
        * 38295 38394: gap of 100 bp
        * 38395 116218: contig of 77824 bp in length
        * 116219 116318: gap of 100 bp
        * 116319 119531: contig of 3213 bp in length
        * 119532 119631: gap of 100 bp
        * 119632 155280: contig of 35649 bp in length
        * 155281 155380: gap of 100 bp
        * 155381 159571: contig of 4191 bp in length
        * 159572 159671: gap of 100 bp
        * 159672 162325: contig of 2654 bp in length.
        * Location/Qualifiers
FEATURES

```

```

source      1. .162325
             /organism="Danio rerio"
             /mol_type="genomic DNA"
             /db_xref="taxon:7955"
             /clone="DKEX-73P2"
             /clone_lib="DanioKey"
misc_feature 1. .17481
             /note="assembly fragment:00118
             fragment_chain:1"
             17582. .38294
             /note="assembly fragment:00360
             fragment_chain:1"
             38395. .116218
             /note="assembly fragment:01043
             fragment_chain:1"
             116319. .119531
             /note="assembly fragment:00044
             fragment_chain:1"
             119632. .155280
             /note="assembly fragment:00616
             fragment_chain:1"
             155381. .159571
             /note="assembly fragment:00075
             fragment_chain:1"
             159672. .162325
             /note="assembly fragment:00014.0"
ORIGIN
Query Match      70.7%; Score 21.2; DB 2; Length 162325;
Best Local Similarity 88.5%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      4 AATGTATGTCGGGTGACATCTATGA 29
      |||||
Db      100707 AATGTGTCGGGTGTCATCTGTA 100732

RESULT 11
AC110515
LOCUS      AC110515      213033 bp      DNA      linear      HTG 09-MAR-2004
DEFINITION Mus musculus chromosome 13 clone RP23-295C1 map 13, WORKING DRAFT
SEQUENCE, 8 ordered pieces.
AC110515
VERSION      AC110515.4      GI:44886734
KEYWORDS      HTG; HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 213033)
AUTHORS      Birren,B., Nusbaum,C. and Lander,E.
TITLE      Mus musculus chromosome 13, clone RP23-295C1
REFERENCE      2 (bases 1 to 213033)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,l., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Farreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McSwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (13-FEB-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 213033)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 10790: contig of 10790 bp in length
* 10791 10890: gap of 100 bp
* 10891 66637: contig of 55747 bp in length
* 66638 66737: gap of 100 bp
* 66738 69878: contig of 3141 bp in length
* 69879 69978: gap of 100 bp
* 69979 77514: contig of 7536 bp in length
* 77515 77614: gap of 100 bp
* 77615 88730: contig of 11116 bp in length
* 88731 88830: gap of 100 bp

```

* 88831 92930: contig of 4100 bp in length
 * 92931 93030: gap of 100 bp
 * 93031 188844: contig of 95814 bp in length
 * 188845 188944: gap of 100 bp
 * 188945 213033: contig of 24089 bp in length.

FEATURES

source
 1..213033
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="13"
 /map="13"
 /clone="RP23-295C1"
 /clone_lib="RPC1-23 Female Mouse BAC"
 1..10790
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 10891..66637
 /note="assembly_fragment"
 66738..69878
 /note="assembly_fragment"
 69979..77514
 /note="assembly_fragment"
 77615..88730
 /note="assembly_fragment"
 88831..92930
 /note="assembly_fragment"
 93031..188844
 /note="assembly_fragment"
 188945..213033
 /note="assembly_fragment
 clone_end:T7
 vector_side:right"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Query Match 70.7%; Score 21.2; DB 2; Length 213033;
 Best Local Similarity 88.5%; Pred. No. 24;
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 CAATGATATGTCGGGTACATCTATG 28
 Db 9398 CTATGATATGTCGGATGACGTCTATG 9423

RESULT 12

AC146937/c
 LOCUS 140142 bp DNA linear PLN 19-JUN-2004
 DEFINITION Oryza sativa chromosome 11 BAC clone OSJNBa0004015, complete sequence.

ACCESSION

AC146937 GI:48958697

VERSION

KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoaceae; Oryza.
 1 (bases 1 to 140142)
 Buell,C.R., Yuan,Q., Ouyang,S., Liu,J., Gansberger,K., Jones,K.M.,
 Overton II,L., Tsirlin,T., Kim,M., Bera,J., Jin,S., Fadrosch,D.W.,
 Tallon,L., Koo,H., Zismann,V., Heiao,J., Blunt,S., Vanaken,S.,
 Riedmuller,S.B., Uterbach,T., Feldblyum,T., Yang,Q., Haas,B.,
 Sub,B., Peterson,J., Quackenbush,J., White,O., Salzberg,S. and
 Fraser,C.

REFERENCE

1 Oryza sativa ssp. japonica cv. Nipponbare OSJNBa0004015 BAC genomic sequence
 Unpublished
 2 (bases 1 to 140142)
 Buell,R.

TITLE

JOURNAL

Submitted (25-OCT-2003) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE

3 (bases 1 to 140142)
 Buell,R.
 Direct Submission
 Submitted (30-OCT-2003) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 13

AC025586

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 13

AC025586

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 13

AC025586

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 13

AC025586

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 13

AC025586

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 13

AC025586

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 13

AC025586

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 13

AC025586

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 13

AC025586

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

ORIGIN

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 13

AC025586

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-320C8"
/clone_lib="RPCI-23"
complement(23305..23539)
/note="We believe the assembly to be correct. The
sequence is a simple repeat (TCCC)n in which the exact
number of repeat copies is unknown. One high quality
subclone in the region spans the repeat into unique
sequence on both sides."

misc_feature
ORIGIN
Query Match 70.0%; Score 21; DB 10; Length 185386;
Best Local Similarity 82.8%; Pred. No. 30;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACATGTATGTCGGGTGATCATCTATGA 29
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98187 AACAACTGAAGTCTGGTTTAGATCTAAGA 98215

RESULT 14
AL592422 205668 bp DNA linear ROD 16-FEB-2002
LOCUS Mouse DNA sequence from clone RP23-218016 on chromosome 11,
DEFINITION complete sequence.
ACCESSION AL592422
VERSION AL592422.11 GI:188955205
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 205668)
Almeida,J.
REFERENCE
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 21, 2002 this sequence version replaced gi:17426517.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-218016 is
from the RPCI-23 Mouse PAC Library
constructed by the group of Pieter de Jong.
For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP23-218016.
FEATURES
source Location/Qualifiers
1..205668
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="11"
/clone="RP23-218016"
/clone_lib="RPCI-23"
108825..108843

misc_feature

/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-320C8"
/clone_lib="RPCI-23"
complement(23305..23539)
/note="We believe the assembly to be correct. The
sequence is a simple repeat (TCCC)n in which the exact
number of repeat copies is unknown. One high quality
subclone in the region spans the repeat into unique
sequence on both sides."

misc_feature
ORIGIN
Query Match 70.0%; Score 21; DB 10; Length 205668;
Best Local Similarity 82.8%; Pred. No. 30;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AACATGTATGTCGGGTGATCATCTATGA 29
||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 183167 AACAACTGAAGTCTGGTTTAGATCTAAGA 183195

RESULT 15
AL136130 69674 bp DNA linear PRI 04-MAR-2003
LOCUS Human DNA sequence from clone RP1-39M18 on chromosome 6q26-27,
DEFINITION complete sequence.
ACCESSION AL136130
VERSION AL136130.7 GI:6911361
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 69674)
Laird,G.
REFERENCE
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 7, 2000 this sequence version replaced gi:5782223.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
RP1-39M18 is from the library RPCI-1 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm

```

```

FEATURES             VECTOR: pCYPAC2.
  source              Location/Qualifiers
    1..69674
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="6"
      /map="Q26-27"
      /clone="RP1-39M18"
      /clone_lib="RPC1-1"

ORIGIN
Query Match          68.7%; Score 20.6; DB 9; Length 69674;
Best Local Similarity 85.2%; Pred. No. 44;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 ACATGTATGTCGGGTGTACATCTATG 28
    |||||
Db 8130 AAACGTATGTCAGTGTACATTTATG 8156

RESULT 16
AC023516              78026 bp   DNA   linear   HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-11504, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC023516
VERSION AC023516.2 GI:9112723
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 78026)
REFERENCE 1
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,B.
TITLE Homo sapiens, clone RP11-11504
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 78026)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,B., Abraham,H., Allen,N.,
  Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
  Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
  Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
  DeArrellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
  Feneator,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
  Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
  Grand-Pierre,N., Grant,G., Hacos,B., Heaford,A., Horton,L.,
  Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
  Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
  Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
  McEwan,P., McEwan,K., McKernan,K., McPheeters,R., Meldrim,J.,
  Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J., Naylor,J.,
  Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
  Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
  Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
  Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
  Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
  Travers,M., Triglio,J., Vassiliev,H., Viel,R., Vo.A., Wilson,B.,
  Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
  Zody,M.
Direct Submission
Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6978142.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Center: Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L1909
Center clone name: 115_O_4

```

* NOTE: This record contains 84 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 800: contig of 800 bp in length
 801 900: gap of 100 bp
 901 1749: contig of 849 bp in length
 1750 1849: gap of 100 bp
 1850 2673: contig of 824 bp in length
 2674 2773: gap of 100 bp
 2774 3614: contig of 841 bp in length
 3615 3714: gap of 100 bp
 3715 4522: contig of 808 bp in length
 4523 4622: gap of 100 bp
 4623 5455: contig of 833 bp in length
 5456 5555: gap of 100 bp
 5556 6388: contig of 833 bp in length
 6389 6488: gap of 100 bp
 6489 7307: contig of 819 bp in length
 7308 7407: gap of 100 bp
 7408 8226: contig of 819 bp in length
 8227 9160: contig of 834 bp in length
 9161 9260: gap of 100 bp
 9261 10064: contig of 804 bp in length
 10065 10164: gap of 100 bp
 10165 10996: contig of 832 bp in length
 10997 11097: gap of 100 bp
 11097 12010: contig of 814 bp in length
 12011 12855: contig of 845 bp in length
 12856 12955: gap of 100 bp
 12956 13800: contig of 845 bp in length
 13801 13900: gap of 100 bp
 13901 14730: contig of 830 bp in length
 14731 14830: gap of 100 bp
 14831 15675: contig of 845 bp in length
 15676 15775: gap of 100 bp
 15776 16610: contig of 835 bp in length
 16611 16710: gap of 100 bp
 16711 17539: contig of 829 bp in length
 17540 17639: gap of 100 bp
 17640 18474: contig of 835 bp in length
 18475 18574: gap of 100 bp
 18575 19389: contig of 815 bp in length
 19390 19489: gap of 100 bp
 19490 20328: contig of 839 bp in length
 20329 20428: gap of 100 bp
 20429 21376: contig of 848 bp in length
 21377 21775: gap of 100 bp
 21776 22174: contig of 798 bp in length
 22175 22274: gap of 100 bp
 22275 23113: contig of 839 bp in length
 23114 23213: gap of 100 bp
 23214 24058: contig of 844 bp in length
 24059 24157: gap of 100 bp
 24158 25006: contig of 849 bp in length
 25007 25106: gap of 100 bp
 25107 25935: contig of 829 bp in length
 25936 26035: gap of 100 bp
 26036 26870: contig of 835 bp in length
 26871 27812: gap of 100 bp
 27813 27912: contig of 842 bp in length
 27913 28735: contig of 823 bp in length

AC073081/c
LOCUS AC073081 118001 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-310H21 from 2, complete sequence.
ACCESSION AC073081
VERSION AC073081.5 GI:16077041
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston, J.E. and Waterston, R.
1 (bases 1 to 118001)
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
9847074
2 (bases 1 to 118001)
Nguyen, C., Kozlowski, A., Boyer, E. and Bielicki, L.
The sequence of Homo sapiens BAC clone RP11-310H21
Unpublished (2001)
3 (bases 1 to 118001)
Waterston, R.H.
Direct Submission
Submitted (08-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 118001)
Waterston, R.H.
Direct Submission
Submitted (12-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 118001)
Waterston, R.
Direct Submission
Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 12, 2001 this sequence version replaced gi:15528929.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
----- Center project name: H_NH0310H21

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tareno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.reagen.com>) or Pieter de Jong

and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-406M18, 2000 bp overlap.
Actual end of this clone is at base position 118001 of RP11-310H21.

Data from AC092430 was used to finish this clone, AC073081.
Polymorphisms have been identified between AC073081 and AC092430. A PCR only Location/Qualifiers

FEATURES	source
1..118001	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-310H21"
	/clone_lib="RPCI-11"
1142..1164	/rpt_family="(T)n"
2410..2463	/rpt_family="(TG)n"
4518..4543	/rpt_family="AT-rich"
5397..6089	/rpt_family="L1"
6215..7134	/rpt_family="L1"
7149..7775	/rpt_family="ERVL"
14542..14752	/rpt_family="L1"
15940..16121	/rpt_family="MIR"
16109..16235	/rpt_family="L2"
16495..16638	/rpt_family="(TA)n"
16689..16723	/rpt_family="AT-rich"
16717..16820	/rpt_family="MIR"
17385..17462	/rpt_family="L2"
17944..18214	/rpt_family="Alu"
21416..21480	/rpt_family="L2"
21720..21816	/rpt_family="MER1_type"
22564..22642	/rpt_family="L2"
22859..23101	/rpt_family="MIR"
23509..23797	/rpt_family="Alu"
24221..24359	/rpt_family="MIR"
24410..24442	/rpt_family="AT-rich"
24962..25144	/rpt_family="ERVL"
25149..25203	/rpt_family="ERVL"
25189..25369	/rpt_family="(TA)n"
25381..25433	/rpt_family="(TA)n"
25429..25927	/rpt_family="ERVL"
25955..26078	/rpt_family="ERVL"

```

repeat_region 26220. .27199
/rpt_family="ERV1"
repeat_region 27200. .27490
/rpt_family="Alu"
repeat_region 27484. .27521
/rpt_family=" (GGA) n"
repeat_region 27553. .27610
/rpt_family=" (GAAA) n"
repeat_region 27625. .28477
/rpt_family="ERV1"
repeat_region 28478. .28833
/rpt_family="MaLR"
repeat_region 28834. .29073
/rpt_family="ERV1"
repeat_region 29093. .30133
/rpt_family="ERV1"
repeat_region 30199. .30624
/rpt_family="ERV1"
repeat_region 31533. .31758
/rpt_family="L1"
repeat_region 32215. .32616
/rpt_family="MaLR"
repeat_region 33073. .33097
/rpt_family=" (T) n"
repeat_region 33383. .34641
/rpt_family="L1"
repeat_region 33653. .33684
/rpt_family="A-rich"
repeat_region 34087. .34142
/rpt_family="GA-rich"
repeat_region 34230. .34254
/rpt_family="AT-rich"
repeat_region 34774. .35137
/rpt_family="MaLR"
repeat_region 35170. .35535
/rpt_family="L2"
repeat_region 35536. .35822
/rpt_family="Alu"
repeat_region 35823. .35871
/rpt_family="L2"
repeat_region 35872. .36185
/rpt_family="Alu"
repeat_region 36186. .36649
/rpt_family="L2"
repeat_region 36758. .36780
/rpt_family=" (TTCA) n"
repeat_region 36786. .36860
/rpt_family="L2"
repeat_region 39431. .39539
/rpt_family="MIR"
repeat_region 41473. .41782
/rpt_family="L1"
repeat_region 42017. .42054
/rpt_family="AT-rich"
repeat_region 42553. .42856

Query Match      68.0%; Score 20.4; DB 9; Length 118001;
Best Local Similarity 80.0%; Pred. No. 58;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGATCATATGAC 30
||||| ||| ||| ||| ||| ||| ||| |||
Db 81502 AACAAATGACTGGCAGGTGTCAGTATGAC 81473

RESULT 19
AC040996/c
LOCUS AC040996 193910 bp DNA linear HTG 28-MAY-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-347K3 map 4, WORKING DRAFT
SEQUENCE 28 unordered pieces.
ACCESSION AC040996
VERSION AC040996.2 GI:8099895
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (Bases 1 to 193910)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 4, clone RP11-347K3
Unpublished
2 (Bases 1 to 193910)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehocsky, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrim, J., Meneus, J., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, F., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 28, 2000 this sequence version replaced gi:7534198.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9510
Center clone name: 347.K.3
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 179519 bases at least Q40
Consensus quality: 186840 bases at least Q30
Consensus quality: 189577 bases at least Q20
Insert size: 192000; agarose-gel
Insert size: 191210; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; agarose-gel
Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 28 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1164: contig of 1164 bp in length

* 1165 1264: gap of 100 bp

* 1265 1611: contig of 347 bp in length

* 1612 1711: gap of 100 bp

* 1712 3179: contig of 1468 bp in length

* 3180 3279: gap of 100 bp

* 3280 5657: contig of 2378 bp in length

* 5658 5757: gap of 100 bp

```
* 5758 7530: contig of 1773 bp in length
* 7531 7630: gap of 100 bp
* 7631 9525: contig of 1895 bp in length
* 9526 9625: gap of 100 bp
* 9626 12289: contig of 2664 bp in length
* 12290 12389: gap of 100 bp
* 12390 14953: contig of 2564 bp in length
* 14954 15053: gap of 100 bp
* 15054 19484: contig of 4431 bp in length
* 19485 19584: gap of 100 bp
* 19585 22744: contig of 3160 bp in length
* 22745 22844: gap of 100 bp
* 22845 27808: contig of 4964 bp in length
* 27809 33103: contig of 5195 bp in length
* 33104 33203: gap of 100 bp
* 33204 39410: contig of 6207 bp in length
* 39411 39510: gap of 100 bp
* 39511 45603: contig of 6093 bp in length
* 45604 45703: gap of 100 bp
* 45704 49406: contig of 3703 bp in length
* 49407 49506: gap of 100 bp
* 49507 56994: contig of 7488 bp in length
* 56995 57094: gap of 100 bp
* 57095 63379: contig of 6285 bp in length
* 63380 63479: gap of 100 bp
* 63480 71011: contig of 7532 bp in length
* 71012 71111: gap of 100 bp
* 71112 76827: contig of 5716 bp in length
* 76828 76927: gap of 100 bp
* 76928 84775: contig of 7848 bp in length
* 84776 84875: gap of 100 bp
* 84876 90721: contig of 5846 bp in length
* 90722 90821: gap of 100 bp
* 90822 98884: contig of 8063 bp in length
* 98885 98984: gap of 100 bp
* 98985 104479: contig of 5495 bp in length
* 104480 104579: gap of 100 bp
* 104580 115298: contig of 10719 bp in length
* 115299 115398: gap of 100 bp
* 115399 128058: contig of 12660 bp in length
* 128059 128158: gap of 100 bp
* 128159 142924: contig of 14766 bp in length
* 142925 143024: gap of 100 bp
* 143025 163395: contig of 20371 bp in length
* 163396 163495: gap of 100 bp
* 163496 193910: contig of 30415 bp in length.
FEATURES
    source
    1..193910
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="4"
        /map="4"
        /clone="RP11-347K3"
        /clone_lib="RPC1-11 Human Male BAC"
    1..1164
        /note="assembly_fragment"
    1265..1611
        /note="assembly_fragment"
        clone_end:T7
        vector_side:right
    1712..3179
        /note="assembly_fragment"
    3280..5657
        /note="assembly_fragment"
    5758..7530
        /note="assembly_fragment"
    7631..9525
        /note="assembly_fragment"
    9626..12289
        /note="assembly_fragment"
    12390..14953
        /note="assembly_fragment"
```

```
misc_feature /note="assembly_fragment"
15054..19484
misc_feature /note="assembly_fragment"
19585..22744
misc_feature /note="assembly_fragment"
22845..27808
misc_feature /note="assembly_fragment"
27909..33103
misc_feature /note="assembly_fragment"
33204..39410
misc_feature /note="assembly_fragment"
39511..45603
misc_feature /note="assembly_fragment"
45704..49406
misc_feature /note="assembly_fragment"
49507..56994
misc_feature /note="assembly_fragment"
57095..63379
misc_feature /note="assembly_fragment"
63480..71011
misc_feature /note="assembly_fragment"
71112..76827
misc_feature /note="assembly_fragment"
76928..84775
misc_feature /note="assembly_fragment"
84876..90721
misc_feature /note="assembly_fragment"
90822..98884
misc_feature /note="assembly_fragment"
98985..104479
misc_feature /note="assembly_fragment"
104580..115298
misc_feature /note="assembly_fragment"
115399..128058
misc_feature /note="assembly_fragment"
128159..142924
misc_feature /note="assembly_fragment"
143025..163395
misc_feature /note="assembly_fragment"
clone_end:SP6
vector_side:left
163496..193910
misc_feature /note="assembly_fragment"

ORIGIN
Query Match 68.0%; Score 20.4; DB 2; Length 193910;
Best Local Similarity 80.0%; Pred. No. 61;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATATGTCGGGTACATCTATGAC 30
||||| ||||| ||||| ||||| |||||
Db 38306 AACAAATGACTGGCAGGTGTCAGCTATGAC 38277

RESULT 20
AC109822/c
LOCUS AC109822 194235 bp DNA linear PRI 12-JUN-2002
DEFINITION Homo sapiens BAC clone RP11-633J12 from 2, complete sequence.
ACCESSION AC109822
VERSION AC109822.5 GI:21281638
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194235)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 194235)
AUTHORS Grewal,N., Haakenson,W., Dignan,G. and Nguyen,C.
```

The sequence of Homo sapiens BAC clone RP11-633J12
 Unpublished (2001)
 3 (bases 1 to 194235)
 Waterston,R.H.
 Direct Submission
 Submitted (07-FEB-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 194235)
 Waterston,R.H.
 Direct Submission
 Submitted (06-MAR-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 194235)
 Waterston,R.H.
 Direct Submission
 Submitted (31-MAY-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 6 (bases 1 to 194235)
 Waterston,R.H.
 Direct Submission
 Submitted (12-JUN-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On May 31, 2002 this sequence version replaced gi:19172920.

 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@watson.wustl.edu

 Summary Statistics

 Center project name: H_NH0633J12

 NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.
 This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.
 MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D.
 McPherson, Department of Genetics, Washington University, St. Louis
 MO. For additional information about the map position of this
 sequence, see <http://genome.wustl.edu/gsc>
 SOURCE INFORMATION:
 The RPC1-11 human BAC library was made from the blood of one male
 donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
 Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
 approach for construction of bacterial artificial chromosome
 libraries. Genomics 51:1-8. The clone may be obtained either from
 Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong
 and coworkers at <http://www.chori.org>
 VECTOR: pBACE3.6
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the right is RP11-742M10, 2000 bp overlap.
 Actual end is at base position 32235 of RP11-742M10.
 Polymorphisms have been identified between AC092430 and AC109822.
 Data from AC073081 was used to finish this clone, AC109822.
 Location/Qualifiers
 1..194235
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="2"
 /map="2"
 /clone="RP11-633J12"
 /clone_lib="RPC1-11"
 1266..1374
 /rpt_family="MIR"
 3308..3617
 /rpt_family="L1"
 3852..3889
 /rpt_family="AT_rich"
 4388..4695
 /rpt_family="Alu"
 5730..5903
 /rpt_family="L1"
 6317..6415
 /rpt_family="MER1_type"
 6638..6660
 /rpt_family="AT_rich"
 6951..6982
 /rpt_family="AT_rich"
 8353..8546
 /rpt_family="MIR"
 8914..8940
 /rpt_family="Alu"
 9201..9479
 /rpt_family="AT_rich"
 9645..9775
 /rpt_family="L2"
 11441..11667
 /rpt_family="L1"
 11709..11859
 /rpt_family="L1"
 11860..12156
 /rpt_family="Alu"
 12157..12591
 /rpt_family="L1"
 12880..13015
 /rpt_family="L1"
 13760..14284
 /rpt_family="MaLR"
 14842..15205
 /rpt_family="MaLR"
 15592..15932
 /rpt_family="MER1_type"
 16811..16836
 /rpt_family="MIR"
 17542..17710
 /rpt_family="MIR"
 19658..19823
 /rpt_family="MER1_type"
 20081..20158
 /rpt_family="L2"
 21181..21411
 /rpt_family="MaLR"
 21636..21824
 /rpt_family="MER1_type"
 22076..22793
 /rpt_family="L1"
 24725..25022
 /rpt_family="Alu"
 25094..25431
 /rpt_family="(TA)n"
 25432..25462
 /rpt_family="AT_rich"
 25463..25958
 /rpt_family="(TA)n"
 25970..26137
 /rpt_family="(TA)n"
 26149..26313
 /rpt_family="(TA)n"
 26332..26625
 /rpt_family="(TA)n"


```

/note="assembly_name:Contig35"
132824..168406
/note="assembly_name:Contig36"
168507..298385
/note="assembly_name:Contig37
clone_end:SP6
vector_side:right"

Query Match      68.0%; Score 20.4; DB 2; Length 288385;
Best Local Similarity 80.0%; Pred. No. 63;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AACATGATGTCGGGTGTACATCTATGAC 30
Db 253932 AACATGACTGGCAGGTGTCCAGCTATGAC 253903

RESULT 22
AC008034
LOCUS
DEFINITION Homo sapiens chromosome 3 clone CTB-134A23, WORKING DRAFT SEQUENCE,
AC008034
4 unordered pieces.
AC008034 30 GI:22138152
HTG: HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 121101)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,
Alsbrooks,S.L., Anarutunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaris,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Chen,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Caron,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Hollins,B.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.K.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hernandez,J.,
Homsai,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseghe,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,Z., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okunolu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,
Sodergren,E., Sonatke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Ugmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,X.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished

REFERENCE
AUTHORS
Worley,K.C.
Submitted (13-JUN-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 8, 2002 this sequence version replaced gi:20279304.
-----
Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: HMHO
Center clone name: CTB-134A23
-----
Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-primer Bodipy; 33% of reads
Assembly: Dye-terminator Big Dye; 67% of reads
Consensus quality: 119470 bases at least Q40
Consensus quality: 120027 bases at least Q30
Consensus quality: 120436 bases at least Q20
Estimated insert size: 121022; sum-of-ctngs estimation
Quality coverage: 11x in Q20 bases; sum-of-ctngs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* * 1 2831: contig of 2831 bp in length
* * 2832 2931: gap of unknown length
* * 2932 2719: contig of 24288 bp in length
* * 2720 27319: gap of unknown length
* * 27320 81157: contig of 53838 bp in length
* * 81158 81257: gap of unknown length
* * 81258 121101: contig of 39844 bp in length.
* * Location/Qualifiers
* 1. 121101
* /organism="Homo sapiens"
* /mol_type="genomic DNA"
* /db_xref="taxon:9606"
* /chromosome="3"
* /clones="CTB-134A23"

FEATURES
source
Query Match 67.3%; Score 20.2; DB 2; Length 121101;
Best Local Similarity 88.0%; Pred. No. 74;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ATGTATGTCGGGTGTACATCTATGCA 29
Db 93892 ATGTATGTCGAAGGTACATCTGTGA 93816

RESULT 23
AC068315
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-224P21 map 3p, WORKING DRAFT

```

```

SEQUENCE, 28 unordered pieces.
AC068315
VERSION AC068315.4 GI:10312236
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 155313)
Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
Dong,W., Fan,H., Fang,X., Guan,Q., Gu,X., Guo,D., He,L., Hu,S.,
Huang,F., Jin,Y., Kang,N., Li,C., Li,G., Li,J., Li,L., Li,S.,
Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,Y.,
Luo,J., Niu,Y., Qi,Q., Qi,X., Song,S., Sun,M., Sun,W., Sun,Y.,
Tao,R., Wang,H., Wang,J., Wang,L., Wang,L., Wang,R.,
Wang,C., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y.,
Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H.,
Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y.,
Zhang,Z., Zhu,B., Yu,J. and Yang,H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 155313)
Bao,W., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B.,
Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y.,
Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y.,
Li,G., Li,C., Bao,Q., Bao,J., Wang,X., Song,L., Zhang,L., Guo,D.,
Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Li,L., Feng,X., Yu,J.
and Yang,H.
Direct Submission
Submitted (02-MAY-2000) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
On Sep 27, 2000 this sequence version replaced gi:8101156.
-----Genome Center
Center:Beijing Center
Center code:Beijing
Website:http://hgclgtp.ac.cn
http://www.genomics.org.cn
Contact:hgclgtp.ac.cn
----- Project Information
Center project name:1k project
Center Clone name: RP11-224P21
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator; ET 55% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 154843 bases at least Q40
Consensus quality: 165011 bases at least Q30
Consensus quality: 168985 bases at least Q20
Insert size: 139544; sum-of-contigs
Quality coverage: 5.36x in Q20 bases,sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1139: contig of 1139 bp in length
* 1140 1239: gap of unknown length
* 1240 2448: contig of 1209 bp in length
* 2449 2548: gap of unknown length
* 2549 3935: contig of 1387 bp in length
* 3936 4035: gap of unknown length
* 4036 6131: contig of 2096 bp in length
* 6132 6231: gap of unknown length
* 6232 7429: contig of 1198 bp in length
* 7430 7529: gap of unknown length
* 7530 9054: contig of 1525 bp in length
* 9055 9154: gap of unknown length
*
10950: contig of 1796 bp in length
11050: gap of unknown length
12637: contig of 1587 bp in length
12737: gap of unknown length
15121: contig of 2384 bp in length
15221: gap of unknown length
17220: contig of 1999 bp in length
17320: gap of unknown length
20436: contig of 3116 bp in length
20536: gap of unknown length
23761: contig of 3225 bp in length
23861: gap of unknown length
26190: contig of 2329 bp in length
26290: gap of unknown length
28871: contig of 2581 bp in length
28971: gap of unknown length
33062: contig of 4091 bp in length
33162: gap of unknown length
37033: contig of 3871 bp in length
37133: gap of unknown length
43465: contig of 6332 bp in length
43565: gap of unknown length
48805: contig of 5240 bp in length
48905: gap of unknown length
56158: contig of 7253 bp in length
56258: gap of unknown length
63409: contig of 7151 bp in length
63509: gap of unknown length
72539: contig of 9030 bp in length
72639: gap of unknown length
83389: contig of 10750 bp in length
83489: gap of unknown length
91426: contig of 7937 bp in length
91526: gap of unknown length
99509: contig of 7983 bp in length
99609: gap of unknown length
108238: contig of 8629 bp in length
108338: gap of unknown length
119079: contig of 10741 bp in length
119179: gap of unknown length
135925: contig of 16746 bp in length
136025: gap of unknown length
155313: contig of 19288 bp in length.
FEATURES
source
1. 155313
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p"
/clone="RP11-224P21"
misc_feature
1. 1139
/note="assembly_name:Contig32"
1240. 2448
/note="assembly_name:Contig33"
2549. 3935
/note="assembly_name:Contig34"
4036. 6131
/note="assembly_name:Contig35"
6232. 7429
/note="assembly_name:Contig36"
7530. 9054
/note="assembly_name:Contig37"
9155. 10950
/note="assembly_name:Contig38"
11051. 12637
/note="assembly_name:Contig39"
12738. 15121
/note="assembly_name:Contig40"
15222. 17220
/note="assembly_name:Contig41"
17321. 20436
/note="assembly_name:Contig42"

```

```

misc_feature 20537..23761
/note="assembly_name:Contig43"
misc_feature 23862..26190
/note="assembly_name:Contig44"
misc_feature 26291..28871
/note="assembly_name:Contig45"
misc_feature 28972..33062
/note="assembly_name:Contig46"
misc_feature 33163..37033
/note="assembly_name:Contig47"
misc_feature 37134..43465
/note="assembly_name:Contig48"
misc_feature 43566..48805
/note="assembly_name:Contig49"
misc_feature 48906..56158
/note="assembly_name:Contig50"
misc_feature 56259..63409
/note="assembly_name:Contig51"
misc_feature 63510..72539
/note="assembly_name:Contig52
vector_end:SP6
72640..83389
/note="assembly_name:Contig53"
misc_feature 83490..91426
/note="assembly_name:Contig54"
misc_feature 91527..99509
/note="assembly_name:Contig55"
misc_feature 99610..108238
/note="assembly_name:Contig56"
misc_feature 108339..119079
/note="assembly_name:Contig57"
misc_feature 119180..135925
/note="assembly_name:Contig58"
misc_feature 136026..155313
/note="assembly_name:Contig59"

ORIGIN
Query Match 67.3%; Score 20.2; DB 2; Length 155313;
Best Local Similarity 88.0%; Pred. No. 75;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTACATCTATGA 29
|||||
Db 42000 ATGTATGTCAGGTACATCTGTGA 42024

RESULT 24
AC066599 157007 bp DNA linear HTG 19-SBP-2000
LOCUS Homo sapiens chromosome 3 clone RP11-596J9 map 3p, WORKING DRAFT
DEFINITION AC066599 SEQUENCE, 14 unordered pieces.
ACCESSION AC066599.3 GI:10190763
VERSION HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157007)
Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
Dong,W., Fan,H., Feng,X., Guan,Q., Gu,X., Guo,D., He,B., Hu,S.,
Huang,F., Jin,Y., Kang,N., Li,C., Li,G., Li,J., Li,L.,
Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,X.,
Luo,J., Niu,Y., Qi,Q., Qi,X., Song,S., Sun,M., Sun,W., Sun,Y.,
Tao,R., Wang,H., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R.,
Wang,X., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y.,
Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H.,
Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,X., Zhang,X.,
Zhang,Z., Zhu,B., Yu,J. and Yang,H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 157007)

TITLE
JOURNAL
REFERENCE
```

AUTHORS

Wang,J., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B.,
 Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y.,
 Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y.,
 Li,G., Li,C., Bao,Q., Bao,J., Wang,X., Song,L., Zhang,L., Guo,D.,
 Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Li,L., Feng,X., Yu,J.
 and Yang,H.
 Direct Submission
 Submitted (25-APR-2000) Human Genomic Center, Institute of
 Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
 100101, P.R.China
 On Sep 19, 2000 this sequence version replaced gi:8101249.

COMMENT

-----Genome Center
 Center:Beijing Center
 Center code:Beijing
 Website:http://hgsc.igtp.ac.cn
 http://www.genomics.org.cn
 Contact:hgsc@igtp.ac.cn
 ----- Project Information
 Center project name:1# project
 Center clone name: RP11-596J9
 ----- Summary Statistics
 Sequencing vector: pUC18; 100% of reads
 Chemistry: Dye-terminator: ET 55% of reads
 Chemistry: Dye-terminator Big Dye; 45% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 150014 bases at least Q40
 Consensus quality: 153515 bases at least Q30
 Consensus quality: 155718 bases at least Q20
 Insert size: 149003; sum-of-contigs
 Quality coverage: 4.98x in Q20 bases;sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 14 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1641: contig of 1641 bp in length
 1642 1741: gap of unknown length
 1742 5811: contig of 4070 bp in length
 5812 5911: gap of unknown length
 5912 7862: contig of 1951 bp in length
 7863 7962: gap of unknown length
 7963 11614: contig of 3652 bp in length
 11615 11714: gap of unknown length
 11715 15628: contig of 3914 bp in length
 15629 15728: gap of unknown length
 15729 22199: contig of 6471 bp in length
 22200 32786: contig of 10487 bp in length
 32787 32886: gap of unknown length
 32887 39504: contig of 6618 bp in length
 39505 39604: gap of unknown length
 39605 48115: contig of 8511 bp in length
 48116 48215: gap of unknown length
 48216 57788: contig of 9573 bp in length
 57789 57888: gap of unknown length
 57889 69171: contig of 11283 bp in length
 69172 69271: gap of unknown length
 69272 86305: contig of 17034 bp in length
 86306 86405: gap of unknown length
 86406 117887: contig of 31482 bp in length
 117888 117987: gap of unknown length
 117988 157007: contig of 39020 bp in length.

FEATURES

Location/Qualifiers

1..157007
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /map="3p"

source


```

misc_feature      /clone="RP11-596J9"
1. .1641
/note="assembly_name:Contig5"
1742. .5811
/note="assembly_name:Contig6"
5912. .7862
/note="assembly_name:Contig7"
7963. .11614
/note="assembly_name:Contig8"
11715. .15628
/note="assembly_name:Contig9"
15729. .22199
/note="assembly_name:Contig10"
22300. .32786
/note="assembly_name:Contig11"
32887. .39504
/note="assembly_name:Contig12"
39605. .48115
/note="assembly_name:Contig13"
48216. .57788
/note="assembly_name:Contig14"
57889. .69171
/note="assembly_name:Contig15"
69272. .86305
/note="assembly_name:Contig16"
86406. .117887
/note="assembly_name:Contig17"
117988. .157007
/note="assembly_name:Contig18"

```

ORIGIN

```

Query Match      67.3%; Score 20.2; DB 2; Length 157007;
Best Local Similarity 88.0%; Pred. No. 75;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 ATGTATTCGGTCTACATCTATCA 29
||||| ||||| ||||| ||||| |||||
Db 150679 ATGTATTCGAAGTGTACATCTGTGA 150703

```

```

RESULT 25
AC021996/c      161903 bp DNA linear HTG 03-FEB-2000
LOCUS          Homo sapiens chromosome 3p clone RP11-1087020, WORKING DRAFT
DEFINITION     SEQUENCE, 39 unordered pieces.
ACCESSION      AC021996
VERSION        AC021996.1 GI:6742988
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 161903)
AUTHORS        Zeng,Y., Hu,S., Dong,W., Zhang,X., Wang,J., Wang,X., Zhang,Y.,
               Zhang,H., Liu,B., Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X.,
               Cheng,C., Wang,Y., Niu,Y., Qi,X., Li,T., Zhang,H., Liu,N., Wu,D.,
               Yu,B., Fan,H., Liu,Y., Li,G., Li,C., Bao,Q., Bao,J., Wang,X.,
               Song,L., Zhang,L., Guo,D., Huang,F., Zhang,G., Li,J., Bian,X.,
               Zhang,M., Li,L., Feng,X., Yu,J. and Yang,H.
               Chromosome 3p genomic sequence
TITLE          Unpublished
JOURNAL        2 (bases 1 to 161903)
REFERENCE      Zeng,Y., Hu,S., Dong,W., Zhang,X., Wang,J., Wang,X., Zhang,Y.,
               Zhang,H., Liu,B., Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X.,
               Cheng,C., Wang,Y., Niu,Y., Qi,X., Li,T., Zhang,H., Liu,N., Wu,D.,
               Yu,B., Fan,H., Liu,Y., Li,G., Li,C., Bao,Q., Bao,J., Wang,X.,
               Song,L., Zhang,L., Guo,D., Huang,F., Zhang,G., Li,J., Bian,X.,
               Zhang,M., Li,L., Feng,X., Yu,J. and Yang,H.
               Direct Submission
TITLE          Submitted (24-JAN-2000) Human Genomic Center, Institute of
JOURNAL        Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
               100101, P.R.China
COMMENT        * NOTE: This is a 'working draft' sequence. It currently

```

```

* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2162: contig of 2162 bp in length
* 2163 gap of unknown length
* 3554: contig of 1392 bp in length
* 3555 gap of unknown length
* 5163: contig of 1609 bp in length
* 5164 gap of unknown length
* 6571: contig of 1408 bp in length
* 6572 gap of unknown length
* 9023: contig of 2451 bp in length
* 9023 gap of unknown length
* 10618: contig of 1595 bp in length
* 10618 gap of unknown length
* 12743: contig of 2126 bp in length
* 12744 gap of unknown length
* 14960: contig of 2217 bp in length
* 14961 gap of unknown length
* 16933: contig of 1793 bp in length
* 16934 gap of unknown length
* 18727: contig of 1793 bp in length
* 18727 gap of unknown length
* 20268: contig of 1542 bp in length
* 20269 gap of unknown length
* 22024: contig of 1756 bp in length
* 22025 gap of unknown length
* 24217: contig of 2193 bp in length
* 24218 gap of unknown length
* 27012: contig of 2795 bp in length
* 27013 gap of unknown length
* 29737: contig of 2725 bp in length
* 29738 gap of unknown length
* 31314: contig of 1577 bp in length
* 31315 gap of unknown length
* 34024: contig of 2710 bp in length
* 34025 gap of unknown length
* 37280: contig of 3256 bp in length
* 37281 gap of unknown length
* 40667: contig of 3387 bp in length
* 40668 gap of unknown length
* 43987: contig of 3320 bp in length
* 43988 gap of unknown length
* 47756: contig of 3769 bp in length
* 47757 gap of unknown length
* 51701: contig of 3945 bp in length
* 51702 gap of unknown length
* 56619: contig of 4918 bp in length
* 56620 gap of unknown length
* 60240: contig of 3621 bp in length
* 60241 gap of unknown length
* 65975: contig of 5735 bp in length
* 65976 gap of unknown length
* 70348: contig of 4373 bp in length
* 70349 gap of unknown length
* 76646: contig of 8298 bp in length
* 76647 gap of unknown length
* 82485: contig of 5839 bp in length
* 82486 gap of unknown length
* 86968: contig of 4483 bp in length
* 86969 gap of unknown length
* 93334: contig of 6366 bp in length
* 93335 gap of unknown length
* 97637: contig of 4303 bp in length
* 97638 gap of unknown length
* 102998: contig of 5361 bp in length
* 102999 gap of unknown length
* 109894: contig of 6896 bp in length
* gap of unknown length

```

* 109895 116211: contig of 6317 bp in length
 * gap of unknown length
 * 116212 124568: contig of 8357 bp in length
 * gap of unknown length
 * 124569 133483: contig of 8915 bp in length
 * gap of unknown length
 * 133484 140451: contig of 6968 bp in length
 * gap of unknown length
 * 140452 149743: contig of 9292 bp in length
 * gap of unknown length
 * 149744 161903: contig of 12160 bp in length.

FEATURES

source

Location/Qualifiers
 1. .161903
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3p"
 /clone="RP11-1087020"

ORIGIN

Query Match 67.3%; Score 20.2; DB 2; Length 161903;
 Best Local Similarity 88.0%; Pred. No. 76;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ATGTATGTCGGGTGACATCTATGA 29

Db 1257 ATGTATGTCAGGTGACATCTGTGA 1233

RESULT 26

AC022382/c

LOCUS

AC022382 185067 bp DNA linear PRI 13-NOV-2002
 Homo sapiens chromosome 3 clone RP11-266J6 map 3p, complete
 sequence.

ACCESSION

AC022382.4 GI:24942870

VERSION

KEYWORDS

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 185067)

Bao, W., Bao, J., Bao, Q., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,
 Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z.,
 He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F.,
 Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y.,
 Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L.,
 Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,
 Wang, J., Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X.,
 Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X.,
 Yu, B., Zeng, X., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,
 Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,
 Yu, J. and Yang, H.

TITLE

JOURNAL

REFERENCE

AUTHORS

Unpublished
 2 (bases 1 to 185067)
 Wu, D., Hu, S., Dong, W., Zhang, X., Wang, J., Zhang, Y., Zhang, H.,
 Liu, B., Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y.,
 Niu, Y., Qi, X., Li, T., Zhang, H., Liu, N., Wang, X., Yu, B., Fan, H.,
 Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L.,
 Guo, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L.,
 Feng, X., Yu, J. and Yang, H.

Direct Submission

Submitted (03-FEB-2000)

Genetics, Chinese Academy of Sciences, Datun Road, Beijing

100101, P.R.China

3 (bases 1 to 185067)

Bao, W., Bao, J., Bao, Q., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,

Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., He, L.,

Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, G., Li, J.,

Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y., Li, W., Li, W.,

Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M.,

Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J., Wang, J., Wang, L.,

TITLE

JOURNAL

REFERENCE

AUTHORS

Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F.,
 Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B., Zeng, X., Zhang, G.,
 Zhang, H., Zhang, H., Zhang, L., Zhang, M., Zhang, X., Zhang, X.,
 Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N., Yu, J. and Yang, H.
 Direct Submission
 Submitted (02-JAN-2001) Human Genomic Center, Institute of
 Genetics, Chinese Academy of Sciences, Datun Road, Beijing

100101, P.R.China

4 (bases 1 to 185067)

Bao, W., Bao, J., Bao, Q., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,

Dong, W., Fan, H., Feng, X., Gong, J., Guan, Q., Gu, X., Guo, D., Guo, Z.,

He, L., Hu, S., Huang, F., Jin, Y., Kang, N., Li, C., Li, C., Li, F.,

Li, G., Li, J., Li, L., Li, S., Li, T., Liu, Y., Liu, N., Liu, B., Liu, Y.,

Li, W., Li, W., Li, Y., Luo, J., Niu, Y., Qi, Q., Qi, X., Song, L.,

Song, S., Sun, M., Sun, W., Sun, Y., Tan, X., Tao, R., Wang, H., Wang, J.,

Wang, J., Wang, L., Wang, L., Wang, R., Wang, X., Wang, X., Wang, Y.,

Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y., Yan, C., Yang, X., Yu, B.,

Zeng, Y., Zhang, G., Zhang, H., Zhang, H., Zhang, L., Zhang, M.,

Zhang, X., Zhang, X., Zhang, Y., Zhang, Y., Zhang, Z., Zhu, B., Zhu, N.,

Yu, J. and Yang, H.

Direct Submission

Submitted (13-NOV-2002) Human Genomic Center, Institute of

Genetics, Chinese Academy of Sciences, Datun Road, Beijing

100101, P.R.China

On Nov 13, 2002 this sequence version replaced gi:12007691.

-----Genome Center

Center:Beijing Center

Center code:Beijing

Website:http://hgsc.igtp.ac.cn

http://www.genomics.org.cn

Contact:hgsc@igtp.ac.cn

----- Project Information

Center project name:1% project

Center clone name: RP11-266J6

----- Summary Statistics

Sequencing vector: pUC18; 100% of reads

Chemistry: Dye-terminator: ET 55% of reads

Chemistry: Dye-terminator: Big Dye; 45% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 39 bases at least Q40

Consensus quality: 184 bases at least Q30

Consensus quality: 447 bases at least Q20

Insert size: 919; sum-of-contigs

Quality coverage: 0.71x in 920 bases;sum-of-contigs

Location/Qualifiers

1. .185067

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="3"

/map="3p"

/clone="RP11-266J6"

ORIGIN

Query Match

Best Local Similarity 67.3%; Score 20.2; DB 9; Length 185067;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ATGTATGTCGGGTGACATCTATGA 29

Db 113243 ATGTATGTCAGGTGACATCTGTGA 113219

RESULT 27

AC011610/c

LOCUS

AC011610

Homo sapiens chromosome 3 clone RP11-266J6, linear

15 unordered pieces.

ACCESSION

AC011610

VERSION

AC011610.11 GI:11055732

KEYWORDS

HTG; HTGS PHASE1; HTGS_DRAFT.

SOURCE

Homo sapiens (human)

Liu, Y., Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L.,
Feng, D., Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Wang, X.,
Feng, X., Yu, J., and Yang, H.

Chromosome 3p genomic sequence

Unpublished
2 (bases 1 to 197360)
Li, L., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B., Bao, W.,
Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y., Qi, X.,
Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y., Li, G.,
Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D.,
Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Wang, X., Feng, X.,
Yu, J., and Yang, H.

Direct Submission
Submitted (21-DEC-1999) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China

On Feb 3, 2000 this sequence version replaced gi:6684183.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 2243: contig of 2243 bp in length
* gap of unknown length
* 2244 4381: contig of 2138 bp in length
* gap of unknown length
* 4382 6628: contig of 2247 bp in length
* gap of unknown length
* 6629 8467: contig of 1839 bp in length
* gap of unknown length
* 8468 10758: contig of 2291 bp in length
* gap of unknown length
* 10759 13438: contig of 2680 bp in length
* gap of unknown length
* 13439 16301: contig of 2863 bp in length
* gap of unknown length
* 16302 19585: contig of 3284 bp in length
* gap of unknown length
* 19586 23956: contig of 4371 bp in length
* gap of unknown length
* 23957 28296: contig of 4340 bp in length
* gap of unknown length
* 28297 31554: contig of 3258 bp in length
* gap of unknown length
* 31555 35178: contig of 3624 bp in length
* gap of unknown length
* 35179 39632: contig of 4454 bp in length
* gap of unknown length
* 39633 44084: contig of 4452 bp in length
* gap of unknown length
* 44085 48826: contig of 4742 bp in length
* gap of unknown length
* 48827 53193: contig of 4367 bp in length
* gap of unknown length
* 53194 56825: contig of 3632 bp in length
* gap of unknown length
* 56826 61151: contig of 4326 bp in length
* gap of unknown length
* 61152 64920: contig of 3769 bp in length
* gap of unknown length
* 64921 69710: contig of 4790 bp in length
* gap of unknown length
* 69711 75897: contig of 6187 bp in length
* gap of unknown length
* 75898 82172: contig of 6275 bp in length
* gap of unknown length
* 82173 88423: contig of 6257 bp in length
* gap of unknown length
* 88430 95080: contig of 6651 bp in length
* gap of unknown length

* 95081 100966: contig of 5886 bp in length
* gap of unknown length
* 100967 105876: contig of 4910 bp in length
* gap of unknown length
* 105877 113067: contig of 7191 bp in length
* gap of unknown length
* 113068 123225: contig of 10158 bp in length
* gap of unknown length
* 123226 134820: contig of 11595 bp in length
* gap of unknown length
* 134821 149685: contig of 14865 bp in length
* gap of unknown length
* 149686 164537: contig of 14852 bp in length
* gap of unknown length
* 164538 181274: contig of 16737 bp in length
* gap of unknown length
* 181275 197360: contig of 16086 bp in length.

FEATURES

source
Location/Qualifiers
1. 197360
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3p"
/clone="RP11-402p11"

ORIGIN

Query Match 67.3%; Score 20.2; DB 2; Length 197360;
Best Local Similarity 88.0%; Pred. No. 77;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ATGTATGTCGGTGTCATCTATGA 29
|||||

Db 72566 ATGTATGTCGAAGTGTCATCTGTGA 72542
|||||

RESULT 29

BX649535

LOCUS

DEFINITION BX649535 216530 bp DNA linear HTG 10-SEP-2004
Danio rerio clone DKEY-234E12, WORKING DRAFT SEQUENCE, 3 unordered
pieces.

ACCESSION BX649535

VERSION BX649535.6 GI:50234178

KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.

SOURCE

ORGANISM

Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 216530)

AUTHORS

TITLE

JOURNAL

Direct Submission
Submitted (09-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jul 12, 2004 this sequence version replaced gi:50080063.

COMMENT

Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zK234E12
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 213766 bases at least Q40
Consensus quality: 213973 bases at least Q30
Insert size: 216330; sum-of-contigs
Insert size: 165139; 32.2% error; agarose-fp
Quality coverage: 8.57x in Q20 bases; sum-of-contigs Quality
coverage: 11.30x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 2572: contig of 2572 bp in length
 * 2573 2672: gap of 100 bp
 * 2673 214430: contig of 211758 bp in length
 * 214431 214530: gap of 100 bp
 * 214531 216530: contig of 2000 bp in length.

FEATURES

source

1. 216530
 Location/Qualifiers
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone_lib="DREY-234E12"
 /clone_lib="DantioKey"

misc_feature

1. 2572
 /note="assembly_fragment:01340"

misc_feature

2673. 214430
 /note="assembly_fragment:02871"

misc_feature

214531. 216530
 /note="assembly_fragment:02878"

ORIGIN

Query Match 67.3%; Score 20.2; DB 2; Length 216530;
 Best Local Similarity 88.0%; Pred. No. 77;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy

1 AACAACTGATGTCGGGTGATCATCT 25

Db

126617 AACAACTGATGTCGGGTGATCATCT 126641

RESULT 30

AC125770

LOCUS

DEFINITION Rattus norvegicus clone CH230-9K21, *** SEQUENCING IN PROGRESS ***,
 18 unordered pieces.

AC125770

AC125770.2 GI:22773120

VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.

KEYWORDS

Rattus norvegicus (Norway rat)

SOURCE

ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 260792)
 Muzny, D., Marle, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranake, D., Barber, M., Barnstead, M., Benahmed, F., Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyie, M., Cree, A., D'Souza, D., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denison, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, X., Egan, A., Escoto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregregis, E., Geier, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hoques, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, H., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshehwa, L., Loulseghe, H., Lozano, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhiney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Mlosovajevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, D., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartabeyn, A., Slisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission

Unpublished

2 (bases 1 to 260792)

Worley, K. C.

Direct Submission

Submitted (30-JUN-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 260792)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (26-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Sep 10, 2002 this sequence version replaced gi:21630390.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GDHO

Center clone name: CH230-9K21

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 196178 bases at least Q40

Consensus quality: 204206 bases at least Q30

Consensus quality: 208976 bases at least Q20

Estimated insert size: 261938; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 18 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 35223: contig of 35223 bp in length
 * 35224 35223: gap of unknown length
 * 35324 61201: contig of 25878 bp in length
 * 61301: gap of unknown length
 * 61302 88979: contig of 27678 bp in length
 * 88980 92079: gap of unknown length
 * 89080 92598: contig of 3519 bp in length
 * 92599 92698: gap of unknown length
 * 92699 158973: contig of 66275 bp in length
 * 158974 159073: gap of unknown length
 * 159074 185143: contig of 26070 bp in length
 * 185144 185243: gap of unknown length
 * 185244 207716: contig of 22473 bp in length
 * 207717 207816: gap of unknown length
 * 207817 208899: contig of 1083 bp in length
 * 208900 208999: gap of unknown length
 * 209000 210516: contig of 1517 bp in length
 * 210517 210616: gap of unknown length
 * 210617 212412: contig of 1796 bp in length
 * 212413 212512: gap of unknown length
 * 212513 214330: contig of 1918 bp in length
 * 214331 214530: gap of unknown length
 * 214531 218807: contig of 4277 bp in length
 * 218808 218907: gap of unknown length
 * 218908 221762: contig of 2855 bp in length
 * 221763 221862: gap of unknown length
 * 221863 227380: contig of 5518 bp in length
 * 227381 227480: gap of unknown length
 * 227481 230994: contig of 3514 bp in length
 * 230995 231094: gap of unknown length
 * 231095 236110: contig of 5016 bp in length
 * 236111 236210: gap of unknown length
 * 236211 246341: contig of 10131 bp in length
 * 246342 246441: gap of unknown length
 * 246442 260792: contig of 14351 bp in length.

FEATURES

source

1..260792
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-9K21"
 1..2543
 /note="wgs_contig"
 89080..90345
 /note="wgs_contig"
 132723..133897
 /note="wgs_contig"
 145068..147515
 /note="wgs_contig"

ORIGIN

Query Match 67.3%; Score 20.2; DB 2; Length 260792;
 Best Local Similarity 88.0%; Pred. No. 79;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTACATCT 25
 |||||
 Db 27029 AACAAATGATGTCAGGTGTAGCTCT 27053

Search completed: March 18, 2005, 09:27:07
 Job time : 1684 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2005, 06:32:48 ; Search time 252 Seconds
(without alignments)
704.731 Million cell updates/sec

Title: US-09-806-197-24

Perfect score: 30

Sequence: 1 aacaatgtatgtccggtacatcatgac 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : N_Geneseq_16Dec04:*

1: Geneseqn1980s:*

2: Geneseqn1990s:*

3: Geneseqn2000s:*

4: Geneseqn2001as:*

5: Geneseqn2001bs:*

6: Geneseqn2002as:*

7: Geneseqn2002bs:*

8: Geneseqn2003as:*

9: Geneseqn2003bs:*

10: Geneseqn2003cs:*

11: Geneseqn2003ds:*

12: Geneseqn2004as:*

13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	30	3 AAA37981	Aaa37981 PCR prime
2	30	100.0	2048	3 AAA37962	Aaa37962 Plasmid #
3	30	100.0	2048	3 AAA37963	Aaa37963 Plasmid #
4	30	100.0	2052	3 AAA37959	Aaa37959 Carrot pr
5	30	100.0	2052	3 AAA37961	Aaa37961 Carrot pr
6	30	100.0	2055	3 AAA37964	Aaa37964 Plasmid #
7	23.6	78.7	2042	2 AAU15144	AAU15144 New promo
8	19.6	65.3	2931	12 ADN73150	Adn73150 Thale cre
9	19.4	64.7	110000	11 ACN44932	Acn44932 Mouse gen
10	19.2	64.0	42772	11 ACN45116	Acn45116 Mouse gen
11	19	63.3	512	13 ACN51227	Acn51227 Cotton an
12	19	63.3	2700	10 ADB53165	Adb53165 Primary r
13	19	63.3	35042	3 AAA81454	Aaa81454 N. mening
14	19	63.3	110000	3 AAA81489	Continuation (2 of
15	19	63.3	236303	4 AAS11614	Aas11614 Human gen
16	19	63.3	349980	3 AAF21611	Aaf21611 Neisseria
17	18.8	62.7	1630	6 AAS62657	Aas62657 cDNA sequ
18	18.6	62.0	1569	3 AAA30829	Aaa30829 Zebrafish
19	18.6	62.0	1671	3 AAA48445	Aaa48445 Zebrafish
20	18.6	62.0	1671	10 ADC42329	Adc42329 cDNA enco

21	18.6	62.0	2152	3 AAA48446	Aaa48446 Zebrafish
22	18.6	62.0	2152	10 ADC42306	Adc42306 cDNA enco
23	18.6	62.0	2152	10 ADH61060	Adh61060 Zebrafish
24	18.6	62.0	90442	9 ADA03077	Ada03077 Mouse mCG
25	18.6	62.0	90442	9 ADA66361	Ada66361 Mouse mCG
26	18.6	62.0	90442	10 ADB72815	Adb72815 Mouse mCG
27	18.6	62.0	90442	10 ADC26997	Adc26997 Mouse car
28	18.6	62.0	90442	11 ADL27155	Adl27155 Mouse gen
29	18.4	61.3	2501	13 ADR73232	Adr73232 Thale cre
30	18.4	61.3	3177	4 ABL28637	AbL28637 Drosophil
31	18.4	61.3	4354	4 ABL11038	AbL11038 Drosophil
32	18.4	61.3	5460	4 ABL28636	AbL28636 Drosophil
33	18.4	61.3	188017	11 ACN45148	Acn45148 Mouse gen
34	18.2	60.7	4919	4 ABL15204	AbL15204 Drosophil
35	18.2	60.7	40050	9 ADA02585	Ada02585 Mouse Ics
36	18.2	60.7	40050	10 ADB72323	Adb72323 Mouse Ics
37	18.2	60.7	40050	10 ADE95833	Ade95833 Mouse Ics
38	18.2	60.7	310122	13 ABD32533	Abd32533 Mouse can
39	18	60.0	287	12 ADL11418	Adl11418 Cat flea
40	18	60.0	287	12 ADL11392	Adl11392 Cat flea
41	18	60.0	386	12 ADL11486	Adl11486 Cat flea
42	18	60.0	401	4 AAK96285	Aak96285 Human neu
43	18	60.0	401	4 AAK97778	Aak97778 Human neu
44	18	60.0	401	6 ABT01055	Abt01055 Human neu
45	18	60.0	401	6 ABT02548	Abt02548 Human neu
46	18	60.0	448	12 ADL10839	Adl10839 Cat flea
47	18	60.0	83709	12 ADQ19964	Adq19964 Human sof
48	18	60.0	110000	4 AAK95240	Continuation (11 o
49	18	60.0	110000	4 AAK95240	Continuation (12 o
50	18	60.0	110000	4 AAK96733	Continuation (11 o
51	18	60.0	110000	4 AAK96733	Continuation (12 o
52	18	60.0	110000	6 ABT00010	Continuation (11 o
53	18	60.0	110000	6 ABT00010	Continuation (12 o
54	18	60.0	110000	6 ABT01503	Continuation (11 o
55	18	60.0	110000	6 ABT01503	Continuation (12 o
56	18	60.0	110000	12 ADH77486	Continuation (11 o
57	18	60.0	110000	12 ADH77486	Continuation (12 o
58	17.8	59.3	377	4 AAL09572	Aal09572 Human bre
59	17.8	59.3	379	4 AAL17464	Aal17464 Human bre
60	17.8	59.3	428	11 ACN79772	Acn79772 Breast ca
61	17.8	59.3	440	5 AAS67070	Aas67070 DNA encod
62	17.8	59.3	500	9 ACH36742	Ach36742 Human end
63	17.8	59.3	598	4 AAL09479	Aal09479 Human bre
64	17.8	59.3	601	4 AAL17368	Aal17368 Human bre
65	17.8	59.3	662	11 ACN79679	Acn79679 Breast ca
66	17.8	59.3	881	11 ACN89461	Acn89461 Breast ca
67	17.8	59.3	908	5 AAD03340	Aad03340 Probe 8 f
68	17.8	59.3	908	10 ADB91926	Adb91926 Acyltrans
69	17.8	59.3	1425	3 AAC40897	Aac40897 Arabidops
70	17.8	59.3	4338	13 ADS47498	Ads47498 Bacterial
71	17.8	59.3	8893	4 ABL13945	AbL13945 Drosophil
72	17.8	59.3	9739	4 ABL20301	AbL20301 Drosophil
73	17.8	59.3	12979	4 ABL13944	AbL13944 Drosophil
74	17.8	59.3	14727	4 ABL20300	AbL20300 Drosophil
75	17.8	59.3	99014	6 ABN96931	Abn96931 Gene #342
76	17.8	59.3	121129	13 ABD33446	Abd33446 Murine ca
77	17.8	59.3	215126	12 ADQ97362	Adq97362 Mouse can
78	17.6	58.7	201	13 ADS37021	Ads37021 Human aut
79	17.6	58.7	810	3 AAC79975	Aac79975 Human sec
80	17.6	58.7	963	3 AAC47090	Aac47090 Arabidops
81	17.6	58.7	1761	3 AAC51203	Aac51203 Arabidops
82	17.6	58.7	2022	3 AAC50796	Aac50796 Arabidops
83	17.6	58.7	2535	13 ADR07452	Adr07452 Full leng
84	17.6	58.7	4470	12 ADJ39886	Adj39886 Plant cDN
85	17.6	58.7	4508	4 AAL03274	Aal03274 Human rep
86	17.6	58.7	5084	4 AAL03275	Aal03275 Human rep
87	17.6	58.7	5087	4 AAL03276	Aal03276 Human rep
88	17.6	58.7	42339	11 ACN44508	Acn44508 Mouse gen
89	17.6	58.7	94330	11 ACN44662	Acn44662 Human gen
90	17.6	58.7	321019	13 ADS36450	Ads36450 Human aut
91	17.6	58.7	329019	13 ABD32207	Abd32207 Human can
92	17.4	58.0	378	5 AAF67029	Aaf67029 Novel hum
93	17.4	58.0	412	10 ADF66363	Adf66363 Human mic

```

c 94 17.4 58.0 425 5 AAF66854
95 17.4 58.0 452 4 AAI83581
96 17.4 58.0 483 12 ADL11119
c 97 17.4 58.0 663 6 AAS61924
98 17.4 58.0 699 5 AAS67755
c 99 17.4 58.0 1731 5 ABV24103
100 17.4 58.0 1731 5 ABV29984
c 101 17.4 58.0 1813 11 ADO85147
c 102 17.4 58.0 1813 12 ADQ25997
c 103 17.4 58.0 1844 3 AAC48021
c 104 17.4 58.0 2136 11 ADO85146
c 105 17.4 58.0 2136 12 ADQ25996
c 106 17.4 58.0 2318 6 ABL89848
c 107 17.4 58.0 2484 4 AAH16032
c 108 17.4 58.0 2662 2 AAT08977
c 109 17.4 58.0 2676 5 ABV23194
c 110 17.4 58.0 2676 5 ABV22761
c 111 17.4 58.0 2676 5 ABV28586
c 112 17.4 58.0 2676 5 ABV29031
c 113 17.4 58.0 3244 4 ABL08432
c 114 17.4 58.0 3489 12 ADO00910
c 115 17.4 58.0 3545 4 ABL07616
c 116 17.4 58.0 3848 2 AAX27277
c 117 17.4 58.0 4183 6 ABNS9877
c 118 17.4 58.0 4486 10 ADE28256
c 119 17.4 58.0 4519 3 AAC75260
c 120 17.4 58.0 4646 3 AAA49924
c 121 17.4 58.0 4646 10 ACP58216
c 122 17.4 58.0 4890 8 ACC43631
c 123 17.4 58.0 4891 10 ADC30167
c 124 17.4 58.0 4966 10 ADF55434
c 125 17.4 58.0 5365 11 ADL22629
c 126 17.4 58.0 5668 2 AAX27262
c 127 17.4 58.0 5668 4 AAH93872
c 128 17.4 58.0 5668 4 AAS64108
c 129 17.4 58.0 5668 5 ACAS9916
c 130 17.4 58.0 5668 6 ABL95479
c 131 17.4 58.0 5668 8 ACC95643
c 132 17.4 58.0 5668 10 ADB75596
c 133 17.4 58.0 5668 10 ADB14227
c 134 17.4 58.0 5668 10 ACF58211
c 135 17.4 58.0 5668 10 ADG26643
c 136 17.4 58.0 5674 10 ACF58221
c 137 17.4 58.0 5785 8 ABZ36107
c 138 17.4 58.0 5921 8 ACC47326
c 139 17.4 58.0 7522 6 ABL32914
c 140 17.4 58.0 9803 4 AAK79475
c 141 17.4 58.0 12276 4 ABL13674
c 142 17.4 58.0 20001 13 ACN37220
c 143 17.4 58.0 48974 2 AAX55300
c 144 17.4 58.0 49175 13 ABD32540
c 145 17.4 58.0 55235 4 AAK67426
c 146 17.4 58.0 110000 10 ADP77343_10
c 147 17.4 58.0 205388 12 ADQ97560
c 148 17.4 58.0 220224 11 ACN44702
c 149 17.4 58.0 245531 13 ABD33022
c 150 17.4 58.0 290547 13 ABD32598

```

ALIGNMENTS

```

RESULT 1
AAA37981
ID AAA37981 standard; DNA; 30 BP.
XX
AC AAA37981;
XX
DT 18-AUG-2000 (first entry)
XX
DE PCR primer S used for carrot promoter amplification.
XX
KW Carrot; promoter; terminator; transgenic plant; breeding; fertility;

```

```

KW PCR primer; ss.
XX Daucus carota.
OS
XX WO200020613-Al.
XX 13-APR-2000.
XX
XX 28-SEP-1999; 99WO-JP005303.
XX
XX 02-OCT-1998; 98JP-00281124.
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX Nishikawa S, Oeda K;
XX WPI; 2000-303791/26.
XX
XX New Plant promoters and terminators from Daucus carota L., useful in
XX plant breeding, for e.g. controlling fertilities of plants.
XX Example 6; Page 44; 81pp; English.
XX
XX This sequence represents a PCR primer used to amplify a carrot promoter
XX sequence. The invention relates to plant promoters and terminators from
XX Daucus carota L. which are capable of expressing a gene of interest in
XX plants. The invention also includes a chimeric gene characterized in that
XX it comprises the promoter and a desired gene linked to each other in the
XX form capable of functioning. A method of producing a transformant
XX comprises introducing the promoter, the chimeric gene or a vector
XX comprising the promoter and a desired gene or terminator sequence into a
XX host cell. The plant promoters and terminators are useful in plant
XX breeding, for e.g. fertilities of plants may be controlled by expressing,
XX in the host cells, a sense or antisense gene of a male sterility related
XX gene such as S-locus-specific RNase gene
XX
XX Sequence 30 BP; 9 A; 6 C; 6 G; 9 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 30; DB 3; Length 30;
Best Local Similarity 100.0%; Pred. No. 0.00018;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AACAAATGATGTCGGGTGATCATCTATGAC 30
Db 1 AACAAATGATGTCGGGTGATCATCTATGAC 30
RESULT 2
AAA37962
ID AAA37962 standard; DNA; 2048 BP.
XX
XX AAA37962;
AC
XX 18-AUG-2000 (first entry)
DT
XX Plasmid #1 DNA sequence used in mutation of promoter sequence.
DE
XX Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.
XX
XX Daucus carota.
XX
XX WO200020613-Al.
XX 13-APR-2000.
XX
XX 28-SEP-1999; 99WO-JP005303.
XX
XX 02-OCT-1998; 98JP-00281124.
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX Nishikawa S, Oeda K;
XX
XX

```



```

DR WPI; 2000-303791/26.
XX
PT New Plant promoters and terminators from Daucus carota L., useful in
PT plant breeding, for e.g. controlling fertilities of plants.
XX
XX Example 8; Page 71-73; 81pp; English.
XX
CC This sequence represents a plasmid sequence used in a method for
CC introducing a mutation into a carrot promoter. The invention relates to
CC plant promoters and terminators from Daucus carota L. which are capable
CC of expressing a gene of interest in plants. The invention also includes a
CC chimeric gene characterized in that it comprises the promoter and a
CC desired gene linked to each other in the form capable of functioning. A
CC method of producing a transformant comprises introducing the promoter,
CC the chimeric gene or a vector comprising the promoter and a desired
CC terminator sequence into a host cell. The plant promoters and
CC terminators are useful in plant breeding, for e.g. fertilities of plants
CC may be controlled by expressing, in the host cells, a sense or antisense
CC gene of a male sterility related gene such as S-locus-specific RNase gene
XX
XX Sequence 2048 BP; 735 A; 317 C; 316 G; 680 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 30; DB 3; Length 2048;
Best Local Similarity 100.0%; Pred. No. 0.0004;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGATGTCGGGTGTACATCTATGAC 30
Db 1742 AACAAATGATGTCGGGTGTACATCTATGAC 1771
RESULT 4
AAA37959
ID AAA37959 standard; DNA; 2052 BP.
XX
AC AAA37959;
XX
DT 18-AUG-2000 (first entry)
XX
DE Carrot promoter sequence #1.
XX
KW Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.
XX
OS Daucus carota.
XX
PN WO200020613-A1.
XX
PD 13-APR-2000.
XX
PF 28-SEP-1999; 99WO-JP005303.
XX
PR 02-OCT-1998; 98JP-00281124.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Nishikawa S, Oeda K;
XX
WPI; 2000-303791/26.
XX
New Plant promoters and terminators from Daucus carota L., useful in
XX plant breeding, for e.g. controlling fertilities of plants.
XX
XX Claim 1; Page 69-70; 81pp; English.
XX
CC This sequence represents a carrot promoter. The invention relates to
CC plant promoters and terminators from Daucus carota L. which are capable
CC of expressing a gene of interest in plants. The invention also includes a
CC chimeric gene characterized in that it comprises the promoter and a
CC desired gene linked to each other in the form capable of functioning. A
CC method of producing a transformant comprises introducing the promoter,
CC the chimeric gene or a vector comprising the promoter and a desired
CC terminator sequence into a host cell. The plant promoters and
CC terminators are useful in plant breeding, for e.g. fertilities of plants
CC may be controlled by expressing, in the host cells, a sense or antisense
CC gene of a male sterility related gene such as S-locus-specific RNase gene
XX
XX Sequence 2052 BP; 737 A; 317 C; 316 G; 682 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 30; DB 3; Length 2052;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGATGTCGGGTGTACATCTATGAC 30
Db 1746 AACAAATGATGTCGGGTGTACATCTATGAC 1775
RESULT 5
AAA37961
DR WPI; 2000-303791/26.
XX
PT New Plant promoters and terminators from Daucus carota L., useful in
PT plant breeding, for e.g. controlling fertilities of plants.
XX
XX Example 8; Page 71-73; 81pp; English.
XX
CC This sequence represents a plasmid sequence used in a method for
CC introducing a mutation into a carrot promoter. The invention relates to
CC plant promoters and terminators from Daucus carota L. which are capable
CC of expressing a gene of interest in plants. The invention also includes a
CC chimeric gene characterized in that it comprises the promoter and a
CC desired gene linked to each other in the form capable of functioning. A
CC method of producing a transformant comprises introducing the promoter,
CC the chimeric gene or a vector comprising the promoter and a desired
CC terminator sequence into a host cell. The plant promoters and
XX
XX Sequence 2048 BP; 735 A; 317 C; 316 G; 680 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 30; DB 3; Length 2048;
Best Local Similarity 100.0%; Pred. No. 0.0004;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AACAAATGATGTCGGGTGTACATCTATGAC 30
Db 1742 AACAAATGATGTCGGGTGTACATCTATGAC 1771
RESULT 3
AAA37963
ID AAA37963 standard; DNA; 2048 BP.
XX
AC AAA37963;
XX
DT 18-AUG-2000 (first entry)
XX
DE Plasmid #2 DNA sequence used in mutation of promoter sequence.
XX
KW Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.
XX
OS Daucus carota.
XX
PN WO200020613-A1.
XX
PD 13-APR-2000.
XX
PF 28-SEP-1999; 99WO-JP005303.
XX
PR 02-OCT-1998; 98JP-00281124.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Nishikawa S, Oeda K;
XX
WPI; 2000-303791/26.
XX
New Plant promoters and terminators from Daucus carota L., useful in
XX plant breeding, for e.g. controlling fertilities of plants.
XX
XX Example 8; Page 73-74; 81pp; English.
XX
CC This sequence represents a plasmid sequence used in a method for
CC introducing a mutation into a carrot promoter. The invention relates to
CC plant promoters and terminators from Daucus carota L. which are capable
CC of expressing a gene of interest in plants. The invention also includes a
CC chimeric gene characterized in that it comprises the promoter and a
CC desired gene linked to each other in the form capable of functioning. A
CC method of producing a transformant comprises introducing the promoter,
CC the chimeric gene or a vector comprising the promoter and a desired
CC terminator sequence into a host cell. The plant promoters and

```

```

ID AAA37961 standard; DNA; 2052 BP.
XX AC AAA37961;
XX PA 18-AUG-2000 (first entry)
DT XX
DE XX
DE Carrot promoter sequence #2.
XX KW Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.
XX OS Daucus carota.
XX PN WO200020613-A1.
XX PD 13-APR-2000.
XX PF 28-SEP-1999; 99WO-JP005303.
XX PR 02-OCT-1998; 98JP-00281124.
XX PA (SUMO ) SUMITOMO CHEM CO LTD.
XX PI Nishikawa S, Oeda K;
XX DR WPI; 2000-303791/26.
XX PT New Plant promoters and terminators from Daucus carota L., useful in
XX PT plant breeding, for e.g. controlling fertilities of plants.
XX PS Claim 1; Page 78-79; 81pp; English.
XX CC This sequence represents a carrot promoter. The invention relates to
XX CC plant promoters and terminators from Daucus carota L. which are capable
XX CC of expressing a gene of interest in plants. The invention also includes a
XX CC chimeric gene characterized in that it comprises the promoter and a
XX CC desired gene linked to each other in the form capable of functioning. A
XX CC method of producing a transformant comprises introducing the promoter,
XX CC the chimeric gene or a vector comprising the promoter and a desired gene
XX CC terminators are useful in plant breeding, for e.g. fertilities of plants
XX CC may be controlled by expressing, in the host cells, a sense or antisense
XX CC gene of a male sterility related gene such as S-locus-specific RNase gene
XX SQ Sequence 2052 BP; 736 A; 318 C; 317 G; 681 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 3; Length 2052;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGATCATCTATGAC 30
Db 1746 AACAAATGATGTCGGGTGATCATCTATGAC 1775

RESULT 6
AAA37964
ID AAA37964 standard; DNA; 2056 BP.
XX AC AAA37964;
XX PA 18-AUG-2000 (first entry)
DT XX
DE XX
DE Plasmid #3 DNA sequence used in mutation of promoter sequence.
XX KW Carrot; promoter; terminator; transgenic plant; breeding; fertility; ds.
XX OS Daucus carota.
XX PN WO200020613-A1.
XX PD 13-APR-2000.
XX PF 28-SEP-1999; 99WO-JP005303.

Query Match 100.0%; Score 30; DB 3; Length 2052;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGATCATCTATGAC 30
Db 1746 AACAAATGATGTCGGGTGATCATCTATGAC 1775

RESULT 7
AAV15144
ID AAV15144 standard; DNA; 2042 BP.
XX AC AAV15144;
XX PA 02-JUL-1998 (first entry)
DT XX
DE XX
DE New promoter used for root-specific expression in plants.
XX KW Promoter; root; carrot; Kuroda Gosun; root-specific expression;
XX KW plant cell; soil pathogen; improve; nutritive value; edible root plant;
XX KW ss.
XX OS Daucus carota.
XX PN EP824150-A2.
XX PD 18-FEB-1998.
XX PF 12-AUG-1997; 97EP-00113923.
XX PR 12-AUG-1996; 96JP-00212680.
XX PA (SUMO ) SUMITOMO CHEM CO LTD.
XX PI Torikai S, Oeda K;
XX DR WPI; 1998-122310/12.
XX PT New carrot root gene, promoter and terminator - useful in genetic
XX PT engineering for directing root-specific gene expression.
XX PS Claim 2; Page 15-16; 31pp; English.

```

```

XX 02-OCT-1998; 98JP-00281124.
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX Nishikawa S, Oeda K;
XX WPI; 2000-303791/26.
XX New Plant promoters and terminators from Daucus carota L., useful in
XX PT plant breeding, for e.g. controlling fertilities of plants.
XX PS Example 8; Page 74-76; 81pp; English.
XX This sequence represents a plasmid sequence used in a method for
XX CC introducing a mutation into a carrot promoter. The invention relates to
XX CC plant promoters and terminators from Daucus carota L. which are capable
XX CC of expressing a gene of interest in plants. The invention also includes a
XX CC chimeric gene characterized in that it comprises the promoter and a
XX CC desired gene linked to each other in the form capable of functioning. A
XX CC method of producing a transformant comprises introducing the promoter,
XX CC the chimeric gene or a vector comprising the promoter and a desired gene
XX CC terminators are useful in plant breeding, for e.g. fertilities of plants
XX CC may be controlled by expressing, in the host cells, a sense or antisense
XX CC gene of a male sterility related gene such as S-locus-specific RNase gene
XX SQ Sequence 2056 BP; 737 A; 319 C; 318 G; 682 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 3; Length 2056;
Best Local Similarity 100.0%; Pred. No. 0.00041;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGATCATCTATGAC 30
Db 1746 AACAAATGATGTCGGGTGATCATCTATGAC 1775

RESULT 7
AAV15144
ID AAV15144 standard; DNA; 2042 BP.
XX AC AAV15144;
XX PA 02-JUL-1998 (first entry)
DT XX
DE XX
DE New promoter used for root-specific expression in plants.
XX KW Promoter; root; carrot; Kuroda Gosun; root-specific expression;
XX KW plant cell; soil pathogen; improve; nutritive value; edible root plant;
XX KW ss.
XX OS Daucus carota.
XX PN EP824150-A2.
XX PD 18-FEB-1998.
XX PF 12-AUG-1997; 97EP-00113923.
XX PR 12-AUG-1996; 96JP-00212680.
XX PA (SUMO ) SUMITOMO CHEM CO LTD.
XX PI Torikai S, Oeda K;
XX DR WPI; 1998-122310/12.
XX PT New carrot root gene, promoter and terminator - useful in genetic
XX PT engineering for directing root-specific gene expression.
XX PS Claim 2; Page 15-16; 31pp; English.

```


Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AACAAATGATGTCGGGTGATCATCTATGA 29
 |||||
 Db 6156 AACAAATGATGTCAGTGAAACAGGTATTA 6184
 |||||

RESULT 10

ACN45116/C
 ID ACN45116 standard; DNA; 42772 BP.

XX ACN45116;

XX 18-NOV-2004 (first entry)

XX Mouse genomic sequence mC8527.

DE Cytostatic; carcinoma; lymphoma; cancer; murine; gene; ss.

XX Mus musculus.

OS W02003073826-A2.

PN 12-SEP-2003.

XX 28-FEB-2003; 2003WO-US006235.

XX 01-MAR-2002; 2002US-00087192.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-328604/31.

XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 PT comprises a nucleotide sequence.

PS Claim 1; SEQ ID NO 1903; Opp; English.

XX The present invention relates to novel DNA and protein sequences which
 CC are associated with carcinomas. The sequences are useful for: (i) for
 CC screening drug candidates; (ii) for screening of bioactive agent capable
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
 CC carcinoma including lymphoma. The present sequence is one such CA coding
 CC sequence. Note: This patent is an equivalent to basic patent
 CC US2002182586A1, for which no sequence data was published

XX SQ Sequence 42772 BP; 11942 A; 8858 C; 9149 G; 11715 T; 0 U; 1108 Other;

Query Match 64.0%; Score 19.2; DB 11; Length 42772;

Best Local Similarity 87.5%; Pred. No. 1.2e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ATGATGTCGGGTGATCATCTATG 28
 |||||

Db 15706 ATGATGTCGGGTGATCATCTATG 15683
 |||||

RESULT 11

ACN51227
 ID ACN51227 standard; cDNA; 512 BP.

XX ACN51227;

XX 02-DEC-2004 (first entry)

XX

DE

XX

KW

KW

KW

XX

OS

XX

PN

XX

PD

XX

PF

XX

PR

XX

PA

PA

PA

PA

XX

PI

XX

DR

XX

PT

PT

PT

PT

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

XX

SQ

Query Match

Best Local Similarity

Matches

QY

Db

RESULT 11

ID

XX

XX

XX

DT

Cotton androecium tissue EST Clone ID: L1B3828-012-Q1-N6-C3, SEQ:6008.

Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;
 variety Nucotton33B; library L1B3828; molecular tag; molecular marker;
 genetic mapping; molecular mapping; seed germination; plant growth;
 plant quality; plant yield; plant breeding; tissue printing; ss.

Gossypium hirsutum.

US2004123340-A1.

24-JUN-2004.

12-DEC-2001; 2001US-00021323.

14-DEC-2000; 2000US-0255619P.

(DEIK/) DEIKMAN J.

(FENG/) FENG P C C.

(FINC/) FINCHER K L.

(ZIEG/) ZIEGLER T E.

Deikman J, Feng PCC, Fincher KL, Ziegler TE;

WPI; 2004-479808/45.

New isolated nucleic acid molecule that encodes a plant protein or its
 fragment, useful for isolating a variety of agronomically significant
 genes associated with plant growth, quality or yield, and as molecular
 tags to map genes.

Claim 1; SEQ ID NO 6008; 34pp; English.

The invention relates to 17880 cotton expressed sequence tags (ESTs;
 ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
 from primed or non-primed seeds from variety DP50B, mature seeds from
 variety Coker 312 Boswell 96 Field, and androecium tissue, gynoeceum
 tissue, developing fibres, carpel walls and septa from variety
 Nucotton33B. The invention also relates to substantially purified
 proteins or their fragments encoded by nucleic acid molecules of the
 invention, and to transformed plants having a nucleic acid construct
 comprising a nucleic acid of the invention. The cotton ESTs are useful as
 molecular tags to isolate genetic regions, to isolate genes, to map
 genes, to determine gene function and to determine whether genes are
 members of a particular gene family. The nucleic acid molecules may be
 used for isolating a variety of agronomically significant genes
 associated with plant growth, quality, yield and could also serve as
 links in metabolic and catabolic pathways. The nucleic acid molecules are
 also useful for identifying genes important in initiating and maintaining
 seed germination or that may be used to mitigate stresses encountered
 during seed germination. The ESTs additionally enable the acquisition of
 promoters and cis-regulatory elements which will be useful to express
 agronomically significant genes in these tissues and/or other tissues,
 and also permits the acquisition of molecular markers useful in breeding
 schemes, genetic and molecular mapping, and in cloning of agronomically
 significant genes. The nucleic acid molecules are further useful for
 detecting the expression level or pattern of a protein or mRNA and for
 detecting the presence or quantity of a protein by tissue printing. The
 present sequence represents a specifically claimed EST isolated from a
 cotton variety Nucotton33B androecium tissue cDNA library (L1B3828). The
 sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from the US
 patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340

Sequence 512 BP; 159 A; 116 C; 91 G; 146 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 13; Length 512;

Best Local Similarity 81.5%; Pred. No. 63;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 CAATGATGTCGGGTGATCATCTATGA 29
 |||||

|||||

Db 394 CAATCCATTCTGGTGACATCTCTGA 420

RESULT 12

ADBS3615/C

ID ADBS3615 standard; DNA; 2700 BP.

AC ADBS3615;

XX

XX 04-DEC-2003 (first entry)

DT

DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4157.

XX

XX toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;

KW toxicity marker; toxicity progression; drug screening;

KW primary rat hepatocyte toxicity modelling; gene; ds.

XX

OS Rattus norvegicus.

XX

XX WO2003065993-A2.

XX

XX 14-AUG-2003.

XX

XX 04-FEB-2003; 2003WO-US003482.

XX

PR 04-FEB-2002; 2002US-0353171P.

PR 13-MAR-2002; 2002US-0363534P.

PR 08-APR-2002; 2002US-0370248P.

PR 10-APR-2002; 2002US-0371134P.

PR 10-APR-2002; 2002US-0371135P.

PR 10-APR-2002; 2002US-0371150P.

PR 11-APR-2002; 2002US-0371413P.

PR 19-APR-2002; 2002US-0373601P.

PR 19-APR-2002; 2002US-0373602P.

PR 22-APR-2002; 2002US-0374139P.

PR 08-MAY-2002; 2002US-0378370P.

PR 09-MAY-2002; 2002US-0378652P.

PR 09-MAY-2002; 2002US-0378653P.

PR 09-MAY-2002; 2002US-0378665P.

PR 09-JUL-2002; 2002US-0394230P.

PR 09-JUL-2002; 2002US-0394253P.

PR 04-SEP-2002; 2002US-0407688P.

PR 28-JAN-2003; 2003US-0442900P.

XX

(GENE-) GENE LOGIC INC.

PA

PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

PI Elashoff M;

XX

WPI; 2003-731472/69.

XX

XX Determining if a compound induces a toxic effect on a tissue or cell, for

PT identifying hepatotoxic compounds, comprises comparing a gene expression

PT profile of a tissue or cell sample to a database of Tox mean and non-Tox

PT mean values.

XX

XX Claim 44; SEQ ID NO 4157; 874pp; English.

PS

XX

XX The present invention describes a method for determining whether a

CC compound induces a toxic effect on a tissue or cell. The method comprises

CC preparing a gene expression profile of a tissue or cell sample exposed to

CC the compound, and comparing the gene expression profile to a database

CC comprising data or information on the Tox mean and non-Tox mean value.

CC The method is useful for predicting or identifying at least one toxic

CC effect, particularly hepatotoxicity, of a test or unknown compound. The

CC genes listed in the specification are useful as diagnostic or toxicity

CC markers for the prediction or identification of the physiological state

CC of tissue or cell sample that has been exposed to a compound, or to

CC identify or predict the toxic effects of a compound or an agent. These

CC may also be used as markers for monitoring toxicity progression or for

CC drug screening. The present sequence represents a primary rat hepatocyte

CC toxicity modelling related gene sequence from the present invention.

XX

SQ Sequence 2700 BP; 791 A; 633 C; 691 G; 585 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 10; Length 2700;

Best Local Similarity 81.5%; Pred. No. 87;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGTGACATCTATGAC 30

Db 1637 AATGTCTGTCGGTGACATCTCTCAC 1611

RESULT 13

AAA81454

ID AAA81454 standard; DNA; 35042 BP.

XX

XX AAA81454;

XX

DT 04-DEC-2000 (first entry)

DE

DE N. meningitidis partial DNA sequence gnm_2 SEQ ID NO:2.

XX

XX Neisseria meningitidis; Neisseria gonorrhoeae; Genome; immunogenic;

KW antigen; vaccine; diagnosis; infection; antibacterial; identification;

KW Meningococcus B; MenB; ds.

XX

XX Neisseria meningitidis.

OS

XX WO200022430-A2.

XX

PD 20-APR-2000.

XX

XX 08-OCT-1999; 99WO-US023573.

PF

XX 09-OCT-1998; 98US-0103794P.

PR

PR 30-APR-1999; 99US-0132068P.

XX

XX (CHIR) CHIRON CORP.

PA

XX

PI Frazer CW, Hickey E, Peterson J, Tettelin H, Venter JC;

PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

PI Rappuoli R, Pizza M;

XX

XX WPI; 2000-318079/27.

DR

XX

XX Isolated nucleotide sequences of Neisseria meningitidis which can be used

PT in the diagnosis and treatment of N. meningitidis infection and other

PT Neisserial infections, for example, N.gonorrhoea.

XX

PS Claim 7; Page 242-253; 1760pp; English.

XX

XX The present invention describes methods of obtaining immunogenic proteins

CC from Neisseria genomic sequences. AAA81453 to AAA82414 represent

CC specifically claimed Neisseria meningitidis genomic DNA sequences;

CC AAA81260 to AAA81303 and AAB25620 to AAB25663 represent Neisseria DNA

CC sequences and their corresponding proteins; AAA81254 to AAA81259 and

CC AAA81304 to AAA81321 represent PCR primers used in the isolation of

CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent

CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all

CC used in the exemplification of the present invention. The nucleic acid

CC sequences, protein sequences, and antibodies against them, can be used in

CC the manufacture of a composition. The composition can be used as a

CC medicament (or in the manufacture of a medicament) for treating,

CC preventing or diagnosing infection due to Neisserial bacteria. For

CC example, some of the identified proteins could be components of vaccines

CC against Meningococcus B; against all serotypes; and/or against all

CC pathogenic Neisseriae. Identification of sequences from the bacterium

CC will also facilitate production of biological probes, particularly

CC organism-specific probes. Attempts to make efficacious Meningococcus B

CC vaccines have failed mainly due to antigen tolerance. Multivalent

CC vaccines have also been tried but none have successfully overcome

CC antigenic variability. The provision of further, complete sequences may

CC provide an opportunity to identify secreted or surface exposed proteins

CC that may be presumed targets for the immune system and which are not


```

FT      exon      /number= 16
FT      185153. .187765
FT      /*tag= ae
FT      /number= 17
XX      WO200138519-A1.
XX      31-MAY-2001.
XX      24-NOV-2000; 2000WO-AU001435.
XX      26-NOV-1999; 99AU-00004348.
XX      (UYQU ) UNIV QUEENSLAND.
XX      Little M, Yamada T, Holmes G, Georgas K, Kolle G, Wilkinson L;
XX      WPI; 2001-343951/36.
XX      Nucleic acids from human chromosome 2p21-16.3 and the encoded peptide,
XX      useful for preventing, diagnosing and treating e.g. eye disease,
XX      especially cataract formation.
XX      Claim 4; Fig 3; 169pp; English.
XX      The invention relates to nucleic acids from human chromosome 2p21-16.3
XX      and the encoded peptide (and mouse and chicken orthologues) that
XX      comprises a pGECPLP group, an insulin-like growth factor binding protein
XX      (IGFBP)-like domain, cysteine-rich domains, an RGD (undefined) group and
XX      a transmembrane domain. The protein, e.g. CRIM1, interacts with peptides
XX      of the transforming growth factor superfamily. A composition comprising
XX      an expression construct comprising the nucleic acids of the invention or
XX      a mimetic which antagonises or mimics an activity of a CRIM1 polypeptide
XX      may be used in a method for modulating the biological activity of a
XX      polypeptide of the bone morphogenic protein (BMP) family. In this way
XX      they may be used to prevent or treat an eye disease, especially cataract
XX      formation. They may also be used to treat neurodegenerative diseases,
XX      renal and kidney disease, bone and tooth abnormalities, wounds and skin
XX      damage, e.g. by use of the nucleic acid in gene therapy by using
XX      antibodies directed against CRIM1 polypeptides. The present sequence is a
XX      Human genomic DNA containing exons 2-17 of the CRIM1 gene
XX      SQ      Sequence 236303 BP; 70199 A; 46544 C; 47996 G; 71563 T; 0 U; 1 Other;
XX      Query Match      63.3%; Score 19; DB 4; Length 236303;
XX      Best Local Similarity 81.5%; Pred. No. 2.1e+02;
XX      Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX      Qy      4 AATGTATGTCGGGTGATCATCTATGAC 30
XX      Db      133196 AATGTGTGTAGATGTACATCTGTGAC 133222
XX      RESULT 16
XX      AAF21611
XX      ID      AAF21611 standard; DNA; 349980 BP.
XX      AC      AAF21611;
XX      XX      13-MAR-2001 (first entry)
XX      DE      Neisseria meningitidis B nucleotide sequence SEQ ID NO:112.
XX      KW      Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
XX      KW      diagnosis; antigen; detection; infection; gene therapy; antibacterial;
XX      KW      ds.
XX      OS      Neisseria meningitidis.
XX      PN      WO200066791-A1.
XX      PD      09-NOV-2000.
XX      XX

PF      08-MAR-2000; 2000WO-US005928.
XX      30-APR-1999; 99US-0132068P.
XX      08-OCT-1999; 99WO-US023573.
XX      28-FEB-2000; 2000GB-00004695.
XX      (CHIR ) CHIRON CORP.
XX      PA      (GENO-) INST GENOMIC RES.
XX      Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC;
XX      Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
XX      Rappuoli R, Frazer CM, Grandi G;
XX      WPI; 2000-647603/62.
XX      Neisseria meningitidis B full length genome sequence and open reading
XX      frames are used to detect, treat and prevent Neisserial infections.
XX      Claim 7; Appendix A; 692pp; English.
XX      The present invention describes the full length genome of Neisseria
XX      meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
XX      represent fragments of the NMB genomic sequence, as the sequence was too
XX      long to go in a record on its own it was split into 8 sequences which
XX      overlap each other at the beginning and end of each sequence by 4980 bp
XX      (i.e. the last 4980 bp of AAF21544 is repeated at the beginning of
XX      AAF21607, the last 4980 bp of AAF21607 are repeated at the beginning of
XX      AAF21608, and so on). AAF21545 to AAF21588 encode the Neisseria proteins
XX      given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR
XX      primers which are used in the exemplification of the present invention.
XX      The NMB genome and fragments from it have antibacterial activity, and can
XX      be used in vaccines and gene therapy. Neisseria nucleic acids, proteins
XX      and/or antibodies which binds to the proteins can be used in compositions
XX      for treating or preventing infection due to Neisserial bacteria or as a
XX      diagnostic reagent for detecting the presence of Neisserial bacteria or
XX      of antibodies raised to Neisserial bacteria. Computers, computer memory,
XX      computer storage medium or computer databases can be used in a search to
XX      identify open reading frames (ORFs) or coding sequences within the NMB
XX      genome. The DNA sequences provide further opportunities to find antigenic
XX      or immunogenic proteins which are more effective in vaccines than the
XX      outer membrane proteins currently used
XX      SQ      Sequence 349980 BP; 87189 A; 93501 C; 84627 G; 84663 T; 0 U; 0 Other;
XX      Query Match      63.3%; Score 19; DB 3; Length 349980;
XX      Best Local Similarity 81.5%; Pred. No. 2.2e+02;
XX      Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
XX      Qy      1 AACATGATGTCGGGTGATCATCTAT 27
XX      Db      52042 AACATGATTTCTATTGTTCTATCTAT 52068
XX      RESULT 17
XX      AAS62657/C
XX      ID      AAS62657 standard; cDNA; 1630 BP.
XX      AC      AAS62657;
XX      XX      14-FEB-2002 (first entry)
XX      DE      cDNA sequence #444 encoding novel human secreted protein.
XX      KW      Human secreted protein; hyperproliferative disorder; autoimmune disorder;
XX      KW      immune deficiency disorder; blood disorder; inflammatory disorder;
XX      KW      infectious disorder; gene therapy; antimicrobial; hepatotropic;
XX      KW      immunosuppressive; antirheumatic; es.
XX      OS      Homo sapiens.
XX      PN      WO200177291-A2.
XX      PD      18-OCT-2001.
XX      XX

```

PF	28-MAY-1999;	99WO-US011883.
FX		
FR	30-NOV-1998;	98US-0110467P.
XX		
PA	(JUEP//)	JUEPPNER H.
PA	(RUBI//)	RUBIN D A.
XX		
PI	Jueppner H,	Rubin DA;
XX		
DR	WPI; 2000-412319/35.	
DR	P-FSDB; AAY90231.	
XX		
PT	Novel zebrafish parathyroid hormone/parathyroid hormone related peptide receptor 3 and isolated nucleic acid encoding zebrafish parathyroid hormone receptor 1 for treating disorders associated with receptor function.	
PT		
PT		
XX		
PS	Claim 23; Fig 1d; 11lpp; English.	
XX		
CC	This sequence encodes a parathyroid hormone receptor type 3 (PTH3R) receptor protein of the invention. The invention also relates to a PTH3R receptor protein. Antagonists of PTHr or PTH3R can be used for the treatment of diseases associated with an increase in PTHr or PTH3R activity, respectively. The peptides are used for diagnosis or prognosis of diseases and disorders associated with PTH3R or PTHr, such as cancer. The polypeptides can be used as a molecular weight markers on sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gels, or on molecular sieve gel filtration columns. Antigenic epitope-bearing peptides and antibodies, that bind specifically to a polypeptide. The monoclonal antibodies, that bind specifically to a polypeptide. The peptides are useful during diagnosis of diseases and disorders. The CC involving PTHr or PTH3R receptor expression or function. Mutations that affect PTHr or PTH3R sequence and/or expression levels of PTHr or PTH3R could be diagnostic for patients with disease or disorders of a developmental, physiological or neurological nature. The nucleic acid molecules are valuable for chromosome identification. The mapping of DNAs to chromosomes is an important first step in correlating those sequences with genes associated with disease. (Updated on 15-SEP-2003 to standardise OS field)	
CC		
XX		
SQ	Sequence 1569 BP; 382 A; 358 C; 418 G; 411 T; 0 U; 0 Other; Query Match 62.0%; Score 18.6; DB 3; Length 1569; Best Local Similarity 84.0%; Pred. No. 1.2e+02; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
Qy	6 TGTATGTCGGTGACATCTATGC 30 246 TCTGTGCTCTAGTACTATGAC 270	
Db		
RESULT 19		
AAA48445	AA48445	
ID AAA48445	standard; cDNA; 1671 BP.	
XX		
AC	AAA48445;	
XX		
DT	15-SEP-2003 (revised)	
DT	27-OCT-2000 (first entry)	
DE	Zebrafish parathyroid hormone type-3 receptor PTH3R coding sequence.	
KW	Zebrafish; parathyroid hormone type-3 receptor; PTH3R;	
KW	developmental disorder; physiological disorder; neurological disorder;	
XX	ss.	
OS	Danio rerio.	
XX		
Key	Location/Qualifiers	
CDS	97..1671	
FT	/*tag= a	
FT	/product= "PTH3R"	
FT	/partial	


```
XX PN WO200032775-A1.
XX XX
XX PD 08-JUN-2000.
XX PF 30-NOV-1999; 99WO-US028207.
XX PR 30-NOV-1998; 98US-0110467P.
XX PA (JUEP/) JUEPPNER H.
XX PA (RUBI/) RUBIN D A.
XX PI Jueppner H, Rubin DA;
XX XX
XX DR WPI; 2000-412323/35.
XX PT New nucleic acids encoding parathyroid hormone receptors PTH1R and PTH3R,
XX PT useful for treating diseases or disorders associated with impaired
XX PT receptor functions comprises a specific nucleotide sequence.
XX XX
XX PS Disclosure; Fig 1C; 111pp; English.
XX CC The present sequence is the parathyroid hormone type-3 receptor (PTH3R)
XX CC coding sequence from the zebrafish. It was obtained by sequencing a cDNA
XX CC clone. The gene and protein can be used to detect diseases in man where
XX CC the receptor is either overexpressed or underexpressed, and they can be
XX CC used to treat these diseases, which may be developmental, physiological
XX CC or neurological disorders. They can also be used to identify agonists and
XX CC antagonists which can be used in a similar manner. In addition, the gene
XX CC can be used for chromosome identification. (Updated on 15-SEP-2003 to
XX CC standardise OS field)
XX SQ Sequence 1671 BP; 409 A; 381 C; 449 G; 432 T; 0 U; 0 Other;

Query Match 62.0%; Score 18.6; DB 3; Length 1671;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 TGTATGTCGCGTGACATCTATGAC 30
Db 342 TCTGTGCTCTGAGTACATCTATGAC 366

RESULT 20
ID ADC42329 standard; cDNA; 1671 BP.
XX AC ADC42329;
XX 18-DEC-2003 (first entry)
XX DE cDNA encoding zebrafish parathyroid hormone receptor PTH3R seq id 26.
XX XX
XX KW parathyroid hormone; PTH; PTH-related peptide; PthRP;
XX KW parathyroid hormone receptor; PTHR; chromosome identification; zebrafish;
XX KW PTH1R; ss.
XX OS Danio rerio.
XX XX
XX FH Key Location/Qualifiers
XX CDS 97..1668
XX FT /*tag= a
XX FT /product= "zebrafish PTH3R"
XX FT /note= "parathyroid hormone receptor"
XX PN US6541220-B1.
XX XX
XX PD 01-APR-2003.
XX XX
XX PF 30-NOV-1999; 99US-00449632.
XX XX
XX PR 30-NOV-1998; 98US-0110467P.
XX XX

(GEHO ) GEN HOSPITAL CORP.
XX XX
XX PI Jueppner H, Rubin DA;
XX XX
XX DR WPI; 2003-754511/71.
XX DR P-PSDB; ADC42307.
XX XX
XX PT Novel nucleic acid comprising a polynucleotide encoding parathyroid
XX PT hormone/parathyroid hormone-related peptide receptor, useful for
XX PT chromosome identification.
XX PS Disclosure; SEQ ID NO 26; 52pp; English.
XX XX
XX CC The invention describes an isolated nucleic acid (I) comprising a
XX CC polynucleotide having a nucleotide sequence chosen from nucleotide
XX CC sequence encoding a parathyroid hormone (PTH)/PTH-related peptide (PthRP)
XX CC receptor (PTH1R receptor) having a fully defined sequence of 536 amino
XX CC acids as given in the specification, PTH1R receptor, mature PTH1R
XX CC receptor, PTH1R extracellular or transmembrane domain, and their
XX CC complement. (I) is useful for diagnosing and treating decrease in the
XX CC standard or normal level of PTH1R receptor activity in an individual, and
XX CC for chromosome identification. This sequence encodes zebrafish PTH3R.
XX XX
XX SQ Sequence 1671 BP; 409 A; 381 C; 449 G; 432 T; 0 U; 0 Other;

Query Match 62.0%; Score 18.6; DB 10; Length 1671;
Best Local Similarity 84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 TGTATGTCGCGTGACATCTATGAC 30
Db 342 TCTGTGCTCTGAGTACATCTATGAC 366

RESULT 21
AAA48446
ID AAA48446 standard; cDNA; 2152 BP.
XX AC AAA48446;
XX 15-SEP-2003 (revised)
XX DT 27-OCT-2000 (first entry)
XX DE Zebrafish parathyroid hormone type-3 receptor PTH3R coding sequence.
XX XX
XX KW Zebrafish; parathyroid hormone type-3 receptor; PTH3R;
XX KW developmental disorder; physiological disorder; neurological disorder;
XX KW ss.
XX OS Danio rerio.
XX XX
XX FH Key Location/Qualifiers
XX CDS 394..2022
XX FT /*tag= a
XX FT /product= "PTH3R"
XX XX
XX PN WO200032775-A1.
XX XX
XX PD 08-JUN-2000.
XX XX
XX PF 30-NOV-1999; 99WO-US028207.
XX XX
XX PR 30-NOV-1998; 98US-0110467P.
XX XX
XX PA (JUEP/) JUEPPNER H.
XX PA (RUBI/) RUBIN D A.
XX XX
XX PI Jueppner H, Rubin DA;
XX XX
XX DR WPI; 2000-412323/35.
XX DR P-PSDB; AAY99601.
XX XX
XX PT New nucleic acids encoding parathyroid hormone receptors PTH1R and PTH3R,
```

PT useful for treating diseases or disorders associated with impaired
 XX receptor functions comprises a specific nucleotide sequence.
 XX Claim 23; Fig 1D; 111pp; English.
 XX
 CC The present sequence is the parathyroid hormone type-3 receptor (PTH3R)
 CC coding sequence from the zebrafish. It was obtained by sequencing a cDNA
 CC clone. The gene and protein can be used to detect diseases in man where
 CC the receptor is either overexpressed or underexpressed, and they can be
 CC used to treat these diseases, which may be developmental, physiological
 CC or neurological disorders. They can also be used to identify agonists and
 CC antagonists which can be used in a similar manner. In addition, the gene
 CC can be used for chromosome identification. (Updated on 15-SEP-2003 to
 XX standardise OS field)
 XX
 SQ Sequence 2152 BP; 592 A; 471 C; 529 G; 559 T; 0 U; 1 Other;
 Query Match 62.0%; Score 18.6; DB 3; Length 2152;
 Best Local Similarity 84.0%; Pred. No. 1.3e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 6 TGTATGTCGGTGTACATCTATGAC 30
 Db 639 TCTGTCTCCTGAGTACATCTATGAC 663

RESULT 22
 ADC42306
 ID ADC42306 standard; cDNA; 2152 BP.
 AC ADC42306;
 XX
 DT 18-DEC-2003 (first entry)
 DE
 XX cDNA encoding zebrafish parathyroid hormone receptor PTH3R seq id 3.
 XX parathyroid hormone; PTH; PTH-related peptide; PTHrP;
 KW parathyroid hormone receptor; PTHR; chromosome identification; zebrafish;
 KW PTH1R; gene; ss.
 XX
 OS Danio rerio.
 XX
 FH Key Location/Qualifiers
 FT CDS 394..2022
 FT /*tag= a
 FT /product= "Zebrafish PTH3R"
 FT /note= "Parathyroid hormone receptor"
 XX
 PN US6541220-B1.
 XX
 PD 01-APR-2003.
 XX
 PF 30-NOV-1999; 99US-00449632.
 XX
 PR 30-NOV-1999; 98US-0110467P.
 XX
 PA (GEO) GEN HOSPITAL CORP.
 XX
 PI Jueppner H, Rubin DA;
 XX
 DR WPI; 2003-754511/71.
 DR P-PSDB; ADC42307.
 XX
 PT Novel nucleic acid comprising a polynucleotide encoding parathyroid
 PT hormone/parathyroid hormone-related peptide receptor, useful for
 PT chromosome identification.
 XX
 PS Disclosure; SEQ ID NO 3; 52pp; English.
 XX

CC The invention describes an isolated nucleic acid (I) comprising a
 CC polynucleotide having a nucleotide sequence chosen from nucleotide
 CC sequence encoding a parathyroid hormone (PTH)/PTH-related peptide (PTHrP)
 CC receptor (PTH1R receptor) having a fully defined sequence of 536 amino

CC acids as given in the specification, PTH1R receptor, mature PTH1R
 CC receptor, PTH1R extracellular or transmembrane domain, and their
 CC complement. (I) is useful for diagnosing and treating decrease in the
 CC standard or normal level of PTH1R receptor activity in an individual, and
 CC for chromosome identification. This sequence encodes zebrafish PTH3R.
 XX
 SQ Sequence 2152 BP; 592 A; 470 C; 529 G; 560 T; 0 U; 1 Other;
 Query Match 62.0%; Score 18.6; DB 10; Length 2152;
 Best Local Similarity 84.0%; Pred. No. 1.3e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 6 TGTATGTCGGTGTACATCTATGAC 30
 Db 639 TCTGTGTCCTGAGTACATCTATGAC 663

RESULT 23
 ADH61060
 ID ADH61060 standard; cDNA; 2152 BP.
 XX
 AC ADH61060;
 XX
 DT 25-MAR-2004 (first entry)
 DE
 XX Zebrafish PTH3R cDNA from clone zPTH3R.
 XX
 KW Zebrafish; parathyroid hormone; PTH; parathyroid hormone related protein;
 KW PTHrP; parathyroid hormone related protein receptor; PTH1R; PTH3R;
 KW diagnosis; prognosis; pharmaceutical composition; chromosome assay; gene;
 KW ss.
 XX
 OS Danio rerio.
 XX
 FH Key Location/Qualifiers
 FT CDS 394..2022
 FT /*tag= b
 FT /product= "Zebrafish PTH3R protein"
 FT sig_peptide 394..457
 FT /*tag= a
 FT mat_peptide 458..2019
 FT /*tag= c
 FT /product= "Zebrafish mature PTH3R protein"
 XX
 PN US2003162256-A1.
 XX
 PD 28-AUG-2003.
 XX
 PF 25-FEB-2003; 2003US-00372095.
 XX
 PR 30-NOV-1998; 98US-0110467P.
 PR 30-NOV-1999; 99US-00449632.
 XX
 PA (MASS-) MASSACHUSETTS GEN HOSPITAL.
 XX
 PI Jueppner H, Rubin DA;
 XX
 DR WPI; 2003-897927/82.
 DR P-PSDB; ADH61061.
 XX
 PT New parathyroid hormone receptors designated PTH1R and PTH3R isolated
 PT from zebrafish are useful to diagnose and treat parathyroid hormone
 PT receptor-related diseases.
 XX
 PS Claim 23; SEQ ID NO 3; 53pp; English.
 XX
 CC The present invention relates to novel parathyroid hormone (PTH) and
 CC parathyroid hormone related protein (PTHrP) receptors PTH1R and PTH3R
 CC isolated from zebrafish. The invention is useful in the diagnosis and
 CC prognosis of certain diseases and disorders that express significantly
 CC decreased levels of PTH1R and PTH3R. The invention is also useful in
 CC preparing pharmaceutical compositions and in chromosome assays. The
 CC present sequence is zebrafish PTH3R cDNA.

XX SQ Sequence 2152 BP; 592 A; 470 C; 529 G; 560 T; 0 U; 1 Other;
 Query Match 62.0%; Score 18.6; DB 10; Length 2152;
 Best Local Similarity 84.0%; Pred. No. 1.3e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 6 TGTATGTCGGGTACATCTATGAC 30
 Db 639 TCTGTGTCCTGAGTACATCTATGAC 663

RESULT 24
 ADA03077
 ID ADA03077 standard; DNA; 90442 BP.
 XX AC ADA03077;
 XX DT 06-NOV-2003 (first entry)
 XX DE Mouse mCG2257 carcinoma associated gene, SEQ ID NO:1595.
 XX KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;
 KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
 KW gene; ds.
 XX OS Mus sp.
 XX PN WO2003057146-A2.
 XX PD 17-JUL-2003.
 XX PF 26-DEC-2002; 2002WO-US041414.
 XX PR 26-DEC-2001; 2001US-00035832.
 XX PA (SAGR-) SAGRES DISCOVERY.
 XX PI Morris DW;
 XX PI WPI; 2003-587068/55.
 XX DR New recombinant nucleic acid encoding carcinoma associated protein,
 PT useful for preparing compositions for treating carcinomas.
 XX PS Claim 1; SEQ ID NO 1595; 245pp; English.
 XX CC The invention relates to recombinant carcinoma associated (CA) nucleic
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The
 CC invention also encompasses expression vectors and host cells comprising a
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
 CC binds to the protein, and a biochip comprising CA nucleic acid or
 CC fragments thereof. The sequences of the invention were identified using
 CC oncogenic retroviruses, which insert into the genome of the host organism
 CC at random. Many of these do not carry transduced host oncogenes or
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
 CC direct consequence of the effects of proviral integration into host
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
 CC leukaemia) or a propensity to carcinoma by determination of the sequence
 CC of a CA gene, or by determination of CA gene expression in particular
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as
 CC therapeutic agents and in screening and evaluating drug candidates. The
 CC present sequence represents a specifically claimed murine CA nucleic acid
 CC sequence of the invention. Note: The complete sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;
 Query Match 62.0%; Score 18.6; DB 9; Length 90442;
 Best Local Similarity 84.0%; Pred. No. 2.7e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Best Local Similarity 84.0%; Pred. No. 2.7e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTACATCTATGAC 29
 Db 3731 ATGTATGTCGGGCACATATATGA 3755

RESULT 25
 ADA66361
 ID ADA66361 standard; DNA; 90442 BP.
 XX AC ADA66361;
 XX DT 20-NOV-2003 (first entry)
 XX DE Mouse mCG2257 gene genomic DNA sequence.
 XX KW carcinoma-associated gene; CA gene; Rorc gene; mCG15938 gene; BAT1 gene;
 KW Iqgap1 gene; IQGAP1 gene; Zpf29 gene; hCG27579 gene; Kcnj9 gene;
 KW Kcnj9 gene; Pp3cc gene; Pp3CC gene; mCG9110 gene; hCG27579 gene;
 KW cancer cell; lymphatic cell; breast cell; prostate cell; epithelial cell;
 KW carcinoma-associated protein; CAP; cytostatic; gene therapy; anticancer;
 KW vaccine; carcinoma; lymphoma carcinoma; lymphatic cancer; breast cancer;
 KW prostate cancer; DNA vaccine; animal model; mouse; murine; ds; mCG2257.
 XX OS Mus sp.
 XX PN WO2003053224-A2.
 XX PD 03-JUL-2003.
 XX PF 20-DEC-2002; 2002WO-US041776.
 XX PR 20-DEC-2001; 2001US-00034650.
 XX PA (SAGR-) SAGRES DISCOVERY.
 XX PI Morris DW, Engelhard EK;
 XX PI WPI; 2003-569168/53.
 XX DR Novel recombinant carcinoma-associated nucleic acid, useful for
 PT evaluating the effect of a candidate carcinoma drug, and for diagnosing
 XX carcinoma.
 XX PS Claim 1; Page 161-174; 229pp; English.
 XX CC This invention relates to a novel recombinant carcinoma-associated (CA)
 CC nucleic acid comprising a fully defined genomic, mRNA or coding sequences
 CC of mouse Rorc gene or human RORC gene, mouse mCG15938 or human gene BAT1,
 CC mouse Iqgap1 gene or human IQGAP1 gene, mouse Zpf29 gene or human
 CC hCG27579 gene, mouse Kcnj9 gene or human KCNJ9 gene, mouse Pp3cc gene or
 CC human PP3CC gene, mouse mCG9110 gene or human hCG27579 gene, as given in
 CC the specification. CA genes are genes which are preferably expressed in
 CC cancer cells, preferably lymphatic, breast, prostate or epithelial cells.
 CC A compound which modifies the expression of the CA genes or bind to
 CC carcinoma-associated proteins (CAP) may have cytostatic activity and the
 CC sequences of the invention may enable the use of gene therapy or a
 CC development of an anticancer vaccine. Therefore the invention may be
 CC useful for diagnosis and treatment of carcinomas, especially lymphoma
 CC carcinoma, breast cancer and prostate cancer. The CA genes may also be
 CC useful as DNA vaccines and for generating animal models of carcinomas.
 CC The present sequence is that of the mouse mCG2257 gene genomic DNA
 CC sequence of the invention.

XX SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;
 Query Match 62.0%; Score 18.6; DB 9; Length 90442;
 Best Local Similarity 84.0%; Pred. No. 2.7e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTACATCTATGAC 29

```

Db      3731 ATGTATGTCGGGGCACATATATGA 3755
|||||
RESULT 26
ADB72815
ID ADB72815 standard; DNA; 90442 BP.
AC ADB72815;
XX
DT 04-DEC-2003 (first entry)
XX
DE Mouse mCG2257 gene.
XX
KW mouse; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS Mus sp.
XX
FN WO2003008583-A2.
XX
PD 30-JAN-2003.
XX
PF 26-DEC-2001; 2001WO-US051291.
XX
PR 02-MAR-2001; 2001US-00798586.
PR 23-OCT-2001; 2001US-00004113.
PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00937722.
PR 20-DEC-2001; 2001US-00034650.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW, Engelhard EK;
XX
XX WPI; 2003-239337/23.
XX
PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS Claim 1; SEQ ID NO 643; 2304pp; English.
XX
CC The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a mouse gene of the invention.
XX
SQ Sequence 90442 BP; 21913 A; 21998 C; 23663 G; 22868 T; 0 U; 0 Other;
Query Match 62.0%; Score 18.6; DB 10; Length 90442;
Best Local Similarity 84.0%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      5 ATGTATGTCGGGTACATCTATGA 29
|||||
Db      3731 ATGTATGTCGGGGCACATATATGA 3755

RESULT 27
ADC26997
ID ADC26997 standard; DNA; 90442 BP.
XX
AC ADC26997;
XX
DT 18-DEC-2003 (first entry)
XX
DE Mouse carcinoma associated (CA) genomic DNA mCG2257.
XX
KW Carcinoma associated gene; CA gene; PR domain protein 11; PRDM11;
KW carcinoma associated protein; CAP; carcinoma; cancer; adenocarcinoma;

```

```

KW breast cancer; inflammatory carcinoma; Paget's disease; aging;
KW neurodegenerative disease; cytostatic; osteopathic; nootropic;
XX neuroprotective; mouse; ds.
XX Mus sp.
XX US2003087252-A1.
XX 08-MAY-2003.
XX 20-MAR-2002; 2002US-00105637.
XX 22-DEC-2000; 2000US-00747377.
PR 02-MAR-2001; 2001US-00798586.
PR 20-DEC-2001; 2001US-00034650.
XX
PA (MORR/) MORRIS D W.
PA (ENGE/) ENGELHARD E K.
XX
PI Morris DW, Engelhard EK;
XX
XX WPI; 2003-730063/69.
XX
PT Novel recombinant carcinoma-associated gene, PRDM11 useful for diagnosing
PT cancer, and for screening for agents for treating cancers and
PT neurodegenerative diseases.
XX
PS Claim 1; Page 23-35; 49pp; English.
XX
CC The present invention relates to novel carcinoma associated (CA) nucleic
CC acid sequences from both mouse and human. In particular the invention
CC discloses a recombinant nucleic acid comprising a fully defined PR domain
CC protein 11 (PRDM11 - a carcinoma-associated gene) nucleotide sequence.
CC The sequences of the invention are useful for identifying compounds that
CC modulate the activity of a carcinoma associated protein (CAP) which are
CC potentially useful in treating carcinomas. Among the diseases and
CC conditions that may be diagnosed or treated according to the invention
CC are cancers such as adenocarcinoma, breast cancer, inflammatory
CC carcinoma, Paget's disease, etc, aging, and neurodegenerative diseases.
CC The present sequence represents a CA nucleic acid sequence of the
XX invention.
XX
SQ Sequence 90442 BP; 21908 A; 22076 C; 23585 G; 22868 T; 0 U; 5 Other;
Query Match 62.0%; Score 18.6; DB 10; Length 90442;
Best Local Similarity 84.0%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      5 ATGTATGTCGGGTACATCTATGA 29
|||||
Db      3731 ATGTATGTCGGGGCACATATATGA 3755

RESULT 28
ADL27155
ID ADL27155 standard; DNA; 90442 BP.
XX
AC ADL27155;
XX
DT 20-MAY-2004 (first entry)
XX
DE Mouse genomic sequence for mCG2257.
XX
KW Mouse; ds; gene; CA gene; carcinoma associated gene; cytostatic; cancer;
KW carcinoma; lymphoma.
XX
OS Mus sp.
XX
FN US2003216558-A1.
XX
PD 20-NOV-2003.
XX
PF 20-DEC-2001; 2001US-00034650.

```

XX 22-DEC-2000; 2000US-00747377.
 PR 02-MAR-2001; 2001US-00798586.
 XX (MORRIS/) MORRIS D W.
 PA (ENGE/) ENGELHARD E K.
 XX Morris DW, Engelhard EK;
 PI WPI; 2003-902052/82.
 XX WPI; 2003-902052/82.
 XX New carcinoma associated gene, useful for preparing a composition for
 PT diagnosing or treating carcinoma.
 XX Claim 1; SEQ ID NO 43; 342pp; English.
 XX The invention relates to a new recombinant nucleic acid (from a CA,
 CC carcinoma associated, gene) appearing as ADL27113-ADL27172. Also included
 CC are a host cell comprising the recombinant nucleic acid or expression
 CC vector, an expression vector comprising the recombinant nucleic acid, a
 CC recombinant protein (a carcinoma associated protein) comprising the
 CC sequence encoded by the nucleic acid, a method for screening drug
 CC candidates, a method for screening for a bioactive agent capable of
 CC binding to (or modulating the activity of) a carcinoma associated
 CC protein, a method for evaluating the effect of a candidate carcinoma
 CC drug, a method of diagnosing carcinoma, a method for inhibiting the
 CC activity of a carcinoma associated protein, a method of treating
 CC carcinomas, a method of neutralizing the effect of a carcinoma associated
 CC protein, a biochip comprising one or more nucleic acid segments of the
 CC nucleic acid, a method of diagnosing carcinoma or propensity to carcinoma
 CC and a method of determining carcinoma associated gene copy number. The
 CC nucleic acid is useful for preparing a composition for diagnosing or
 CC treating carcinoma especially lymphomas. The present sequence is the
 CC genomic sequence from a mouse carcinoma associated gene.
 XX
 SQ Sequence 90442 BP; 21911 A; 21997 C; 23664 G; 22869 T; 0 U; 1 Other;
 Query Match 62.0%; Score 18.6; DB 11; Length 90442;
 Best Local Similarity 84.0%; Pred. No. 2.7e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 5 ATGTATGTCGGGTACATCTATGA 29
 ||||| ||||| ||||| ||||| |||||
 Db 3731 ATGTATGTCGGGCACATATATGA 3755
 RESULT 29
 ADR73232
 ID ADR73232 standard; DNA; 2501 BP.
 XX ADR73232;
 AC ADR73232;
 XX 16-DEC-2004 (first entry)
 DT Thale cress DNA regulating element (DRE) #6.
 XX Thale cress; DNA regulating element; DRE; ds; plant;
 XX Thale cress; DNA regulating element; DRE; ds; plant;
 KW tissue specific transcription.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX WO2004081173-A2.
 PN WO2004081173-A2.
 XX 23-SEP-2004.
 PD 23-SEP-2004.
 XX 11-MAR-2004; 2004WO-IL000235.
 PF 11-MAR-2004; 2004WO-IL000235.
 XX 12-MAR-2003; 2003US-0453843P.
 PR 12-MAR-2003; 2003US-0453843P.
 XX (EVOG-) EVOGENE LTD.
 PA (EVOG-) EVOGENE LTD.
 XX Karchi H, Meissner R, Ronen G, Golan E, Rabinovich L, Zeligier N;
 PI Savir N;

XX WPI; 2004-668944/65.
 XX New isolated polynucleotide that is a DNA regulating element, capable of
 PT regulating expression of at least one polynucleotide sequence operably
 PT linked to it, useful for directing expression of exogenous polynucleotide
 PT sequences in plants.
 XX Claim 1; SEQ ID NO 26; 131pp; English.
 XX The invention relates to plant derived polynucleotides capable of
 CC regulating expression of at least one polynucleotide sequence operably
 CC linked to it. The invention also relates to a nucleic acid construct
 CC comprising the isolated polynucleotide, a transgenic cell comprising the
 CC isolated polynucleotide or nucleic acid construct, a transgenic organism
 CC comprising the isolated polynucleotide or nucleic acid construct, a
 CC transgenic plant comprising the isolated polynucleotide, a method of
 CC producing a transgenic plant comprising transforming the plant with the
 CC polynucleotide or nucleic acid construct, a method of expressing a
 CC polypeptide of interest in a cell comprising transforming the cell with
 CC the nucleic acid construct thus expressing the polypeptide of interest in
 CC the cell, and a method of co-expressing two polypeptides of interest in a
 CC cell comprising transforming the cell with a nucleic acid construct
 CC including two polynucleotide sequences encoding the two polypeptides of
 CC interest operably linked to a regulatory nucleic acid sequence. The
 CC polynucleotide includes at least one promoter region that is capable of
 CC directing transcription of at least one polynucleotide sequence in a
 CC constitutive, inductive, tissue specific or developmental stage specific
 CC manner. The polynucleotides are useful for directing expression of
 CC exogenous polynucleotide sequences in plants. This sequence represents an
 CC Arabidopsis thaliana DNA regulating element (DRE) used in the scope of
 CC the invention.
 XX Sequence 2501 BP; 807 A; 443 C; 442 G; 809 T; 0 U; 0 Other;
 Query Match 61.3%; Score 18.4; DB 13; Length 2501;
 Best Local Similarity 78.6%; Pred. No. 1.7e+02;
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Oy 1 AACATGTATGTCGGGTACATCTATG 28
 ||||| ||||| ||||| ||||| |||||
 Db 1199 AACTATGTATGTCGGGTACAAATCTTG 1226
 RESULT 30
 ABL28637
 ID ABL28637 standard; DNA; 3177 BP.
 XX ABL28637;
 AC ABL28637;
 XX 26-MAR-2002 (first entry)
 DT Drosophila melanogaster genomic polynucleotide SEQ ID NO 37384.
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 37384.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX Drosophila melanogaster.
 OS Drosophila melanogaster.
 XX WO200171042-A2.
 PN WO200171042-A2.
 XX 27-SEP-2001.
 PD 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US009231.
 PF 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR WPI; 2001-656860/75.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 37384; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB5773-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 3177 BP; 805 A; 848 C; 831 G; 693 T; 0 U; 0 Other;
Query Match 61.3%; Score 18.4; DB 4; Length 3177;
Best Local Similarity 78.6%; Pred. No. 1.7e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 3 CAATGTATGTCGGTGTACATCTATGAC 30
Db 2724 CAATCTGTCCAGCTTACATCTATGCC 2751

Search completed: March 18, 2005, 08:59:08
Job time : 269 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2005, 06:47:48 : Search time 3097 Seconds
(without alignments)
368.721 Million cell updates/sec

Title: US-09-806-197-24
Perfect score: 30
Sequence: 1 aacaatgtatgcgggtacatcatgac 30

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 150 summaries

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_ges1.*
9: gb_ges2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	23.2	77.3	763	8	B11833 F15H3-T7 IG
C 2	21.2	70.7	416	8	A2227906 RPTC1-23-8
C 3	21.2	70.7	773	7	C0958388 AGENCOURT
C 4	21.2	70.7	815	9	AG498011 Mus muscu
C 5	20.8	69.3	631	8	B2666678 PUBCL76TD
C 6	20.8	69.3	977	9	CL989513 ZMMBB0600
C 7	20.6	68.7	631	5	BX507610 DKEZp686A
C 8	20.6	68.7	672	5	BX507786 DKEZp686O
C 9	20.4	68.0	643	6	CA501775 WHB4038 E
C 10	20.4	68.0	707	4	BJ301068
C 11	20.4	68.0	862	9	AG347593 Mus muscu
C 12	20.4	68.0	960	9	CL290684 ZMMBB0603
C 13	20.6	67.7	390	8	BH204000 Sml-54G19
C 14	20.6	67.7	538	9	CL412828 RPTC144_43
C 15	20.6	67.7	542	8	AQ879634 HS 4815 A
C 16	20.6	67.7	605	8	AQ525705 HS_5239 B
C 17	20.6	67.7	773	8	AQ856387 nbe50003A
C 18	20.6	67.7	858	7	CV480979 AGENCOURT
C 19	20.6	67.7	916	8	AZ204484 SP_0056 A
C 20	20.6	67.7	1330	8	B12483 F26A16-T7 I
C 21	19.6	65.3	266	8	BZ091289 CH230-213
C 22	19.6	65.3	380	1	AA751822 96AS0663
C 23	19.6	65.3	385	8	BH121074 RPTC1-24-2
C 24	19.6	65.3	396	7	CF889471 TCTR-272

```

98 19 63.3 709 4 BG545083 602572271
99 19 63.3 723 5 BU459209 603367031
100 19 63.3 743 7 CV113883 AGENCOURT
101 19 63.3 771 7 CN227608 RJ9021E08
102 19 63.3 773 7 CV301588 MM10_C04
103 19 63.3 793 6 CA5111873 UI-R-FJ0-
104 19 63.3 799 9 CG088029 PUJDG61TB
105 19 63.3 816 7 CO070842 GR_Ea28B
106 19 63.3 832 7 CO557970 AGENCOURT
107 19 63.3 839 8 AQ746975 HS_5538_A
108 19 63.3 841 7 CV485419 AGENCOURT
109 19 63.3 847 7 CK148537 AGENCOURT
110 19 63.3 864 9 CG358167 OG3CH19TH
111 19 63.3 876 7 CV483557 AGENCOURT
112 19 63.3 876 9 CW502634 AP_Ba001
113 19 63.3 877 7 CV481130 AGENCOURT
114 19 63.3 899 7 CN201083 TOR1079 G
115 19 63.3 912 8 CC242539 PUHNS77TD
116 19 63.3 917 7 CF265902 AGENCOURT
117 19 63.3 964 9 CG358175 OG3CH19TV
118 19 63.3 976 5 BUL42117 603137058
119 19 63.3 1009 9 CL099762 ISB1-33D8
120 19 63.3 1101 9 CNO500FG Drosophil
121 19 63.3 1156 3 CR683451 Tetraodon
122 19 63.3 1183 3 CR639952 Tetraodon
123 19 63.3 1380 3 CR689919 Tetraodon
124 19 63.3 142 7 CV040804 4138380 B
125 18.8 62.7 302 4 B1314271 dai133R03.
126 18.8 62.7 425 6 CB270918 pgnic.pK0
127 18.8 62.7 439 9 CC998203 ZUAGM63TV
128 18.8 62.7 494 4 BG363783 dc71c02.Y
129 18.8 62.7 536 6 CD470016 Leuko84.1
130 18.8 62.7 544 4 BM180094 dai15e10.
131 18.8 62.7 553 4 B1961434 MONO1.5_E
132 18.8 62.7 570 9 BX973679 Forward_8
133 18.8 62.7 572 4 BM780569 APL1_4_E0
134 18.8 62.7 613 4 BM780480 APL1_3_F0
135 18.8 62.7 616 8 AQ721889 HS_5561_B
136 18.8 62.7 636 9 CC471092 CH240_294
137 18.8 62.7 654 7 CF587673 USPA-FP_1
138 18.8 62.7 668 6 BY760326 BY760326
139 18.8 62.7 729 7 CF290512 AGENCOURT
140 18.8 62.7 730 5 BU254628 603745364
141 18.8 62.7 732 5 BU233627 603409683
142 18.8 62.7 760 9 BX144338 Danilo_xer
143 18.8 62.7 765 4 BG643740 EST511934
144 18.8 62.7 803 9 CG397790 ZMWBc001
145 18.8 62.7 819 1 AU726789 AU726789
146 18.8 62.7 824 5 BU602860 AGENCOURT
147 18.8 62.7 834 5 BU900505 AGENCOURT
148 18.8 62.7 836 9 CG377790 OG3B143TH
149 18.8 62.7 914 6 CA982717 AGENCOURT
150 18.8 62.7 921 6 CA790630 AGENCOURT

```

ALIGNMENTS

```

RESULT 1
B11833/c 763 bp DNA linear GSS 14-MAY-1997
LOCUS F15H3-T7 IGF Arabidopsis thaliana genomic clone F15H3, genomic
DEFINITION survey sequence.
ACCESSION B11833
VERSION B11833.1 GI:2092953
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 763)
AUTHORS Peng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and

```

```

Ecker,J.
BAC End Sequences at ATGC
Unpublished (1997)
Other_GSSs: F15H3-Sp6.1, F15H3-T7.1, F15H3-Sp6
Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA
19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 322
High quality sequence stop: 343.

FEATURES
    source
        1..763
            /organism="Arabidopsis thaliana"
            /mol_type="genomic DNA"
            /ecotype="Columbia"
            /db_xref="taxon:3702"
            /clone="F15H3"
            /sex="hermaphrodite"
            /clone_lib="IGF"
            /notes="Vector: BelOBACII; Site_1: EcoRI; Site_2: EcoRI;
            Produced by Thomas Altmann"

ORIGIN
Query Match 77.3%; Score 23.2; DB 8; Length 763;
Best Local Similarity 89.3%; Pred. No. 8.7;
Matches 25; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACATGTATGTCGGTGTACATCTATCA 29
    |||||||||||||||||||||||||
Db 130 ACGATGTATGTCCTGTGTACCTATGA 103

RESULT 2
LOCUS AZ227906 416 bp DNA linear GSS 14-JUN-2000
DEFINITION RPCI-23-82A18.TJ RPCI-23 Mus musculus genomic clone RPCI-23-82A18,
genomic survey sequence.
ACCESSION AZ227906
VERSION AZ227906.1 GI:8535955
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 416)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
Akinret,B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Krol,M., de
Jong,P. and Fraser,C.M.
Unpublished (1999)
Other_GSSs: RPCI-23-82A18.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_end/mouse/bac_end_intro.html
Plate: 82 row: A column: 18
Seq primer: SP6
Class: BAC ends.

```



```

FEATURES
source
  Location/Qualifiers
    1. .416
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone="RPC1-23-82A18"
      /sex="Female"
      /lab_host="DH10B"
      /clone_lib="RPC1-23"
      /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
      EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
      brain genomic DNA was isolated and partially digested
      with a combination of EcoRI and EcoRI Methylase. Size
      selected DNA was cloned into the pBACe3.6 vector at the
      EcoRI sites. The ligation products were transformed into
      DH10B electrocompetent cells (BRL Life Technologies)."
ORIGIN
  Query Match      70.7%; Score 21.2; DB 8; Length 416;
  Best Local Similarity 88.5%; Pred. No. 67;
  Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 CAATGATATCCGGTGATCATCTATG 28
    | ||||| ||||| ||||| |||||
Db 204 CTATGATATCCGATGACGCTATG 229

RESULT 3
CO958388/c
LOCUS      CO958388      773 bp mRNA linear EST 17-AUG-2004
DEFINITION AGENCOURT 30676684 NIH_ZGC_14 Danio rerio cDNA clone IMAGE:7405493
            5', mRNA sequence.
ACCESSION  CO958388
VERSION     CO958388.1 GI:51322970
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 773)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cga@bse-remail.nih.gov
            Tissue Procurement: John Ngai, Nancy Sugano
            cDNA Library Preparation: Dr. Sumio Sugano
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM15603 row: 1 column: 03
            High quality sequence start: 13
            High quality sequence stop: 557.
            Location/Qualifiers
              1. .773
                /organism="Danio rerio"
                /mol_type="mRNA"
                /db_xref="taxon:7955"
                /clone="IMAGE:7405493"
                /tissue_type="olfactory epithelium"
                /lab_host="DH10B Tona"
                /clone_lib="NIH_ZGC 14"
                /note="Organ: olfactory epithelium; Vector: pME18S-FL3;
                Site1: DraIII; Site2: DraIII; 1st strand cDNA was primed
                with an oligo(dT) primer
                (GCGGCTGAAGACGGCCTATGTCGCCCTTTTTTTTTTTTTTTTTT);
FEATURES
source
  Location/Qualifiers
    1. .773
      /organism="Mus musculus"
      /mol_type="genomic DNA"
      /sub_species="molossinus"
      /db_xref="taxon:57486"
      /clone="MSMg01-397L21.TJ"
      /sex="male"
      /tissue_type="mixture of kidney and spleen"
      /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
  Query Match      70.7%; Score 21.2; DB 9; Length 815;
  Best Local Similarity 88.5%; Pred. No. 75;
  Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

double-stranded cDNA was ligated to a DraIII adaptor
(GGCCACUGG), digested and directionally cloned into
distinct DraIII sites of the pME18S-FL3. Library was size
selected for 1.0 kb, with an average insert size of ~1.2kb.
Library constructed by Yutaka Suzuki (University of Tokyo
Institute of Medical Science). Custom primers recommended
for sequencing: 5' end primer 5'-GGATGTGCGCTTACTTCTA-3'
and 3' end primer 5'-CGACCTGCAGCTCGACACA-3'. Note: This
is a Zebrafish Gene Collection (ZGC) library"
ORIGIN
  Query Match      70.7%; Score 21.2; DB 7; Length 773;
  Best Local Similarity 88.5%; Pred. No. 74;
  Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGGTGTACATCTATGA 29
    | ||||| ||||| ||||| |||||
Db 380 AATGTGTGTCGGGTGTGCACTCTGTA 355

RESULT 4
AG498011
LOCUS      AG498011      815 bp DNA linear GSS 04-JUN-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-397L21.TJ, genomic survey
            sequence.
ACCESSION  AG498011
VERSION     AG498011.1 GI:48205241
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus
ORGANISM    Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS    Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE      BAC end Sequences of Library MSMg01
JOURNAL    Unpublished
            2 (bases 1 to 815)
AUTHORS    Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
            Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institute, Bio Resource Center.
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY      : pBACe3.6
            Vector       : EcoRI
            R.Site 1     : EcoRI.
            R.Site 2     : EcoRI.
            Location/Qualifiers
              1..815
                /organism="Mus musculus molossinus"
                /mol_type="genomic DNA"
                /sub_species="molossinus"
                /db_xref="taxon:57486"
                /clone="MSMg01-397L21.TJ"
                /sex="male"
                /tissue_type="mixture of kidney and spleen"
                /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
  Query Match      70.7%; Score 21.2; DB 9; Length 815;
  Best Local Similarity 88.5%; Pred. No. 75;
  Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 3 CAATGATGTCGGGTGTACATCATG 28
Db 286 CTATGATGTCGGATGTACGCTATG 311

RESULT 5
LOCUS BZ666678 631 bp DNA linear GSS 05-FEB-2003
DEFINITION PUBCL76TD ZM 0.6 1.0 kb Zea mays genomic clone ZMMBtA023N08,
genomic survey sequence.
ACCESSION BZ666678
VERSION BZ666678.1 GI:28212523
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 631)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
ORIGIN
source
Location/Qualifiers
1..631
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBtA023N08"
/clone_lib="ZM_0.6 1.0 kb"
/note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
CoT selected genomic DNA library"

FEATURES
source
Location/Qualifiers
1..631
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBtA023N08"
/clone_lib="ZM_0.6 1.0 kb"
/note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
CoT selected genomic DNA library"

Query Match 69.3%; Score 20.8; DB 8; Length 631;
Best Local Similarity 91.7%; Pred.No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AACATGATGTCGGGTGTACATC 24
Db 458 AACATGATGTCGGGTGCACACC 481

RESULT 6
LOCUS CL989513 977 bp DNA linear GSS 23-SEP-2004
DEFINITION ZMMBHe000609.r ZMMBHe Zea mays genomic clone ZMMBHe000609.3',
genomic survey sequence.
ACCESSION CL989513
VERSION CL989513.1 GI:52557591
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 977)
AUTHORS Ma,J., SanMiguel,P., Liu,R., Haller,K., Soderlund,C. and
Bennetzen,J.
ZMMB sequences
Unpublished (2004)
Contact: Jeff Bennetzen
Bennetzen Lab

Query Match 69.3%; Score 20.8; DB 8; Length 631;
Best Local Similarity 91.7%; Pred.No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AACATGATGTCGGGTGTACATC 24
Db 458 AACATGATGTCGGGTGCACACC 481

RESULT 7
LOCUS BX507610/c 631 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFP686A14261.r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DKFP686A14261 5', mRNA sequence.
ACCESSION BX507610
VERSION BX507610.1 GI:32040422
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 631)
AUTHORS Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
Wiemann,S.
EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
Wellenreuther,R., et al.)
Unpublished (2003)
Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project
No g1 sequence available.
This clone (DKFP686A14261) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1..631
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFP686A14261"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiIB;

FEATURES
source
Location/Qualifiers
1..631
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFP686A14261"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiIB;

The University of Georgia
Department of Genetics, C426a Life Sciences Building, Athens, GA
30602, USA
Tel: 706-542-3698
Fax: 706-583-0972
Email: maize@uga.edu
Plate: 0006 row: 0 column: 09
Class: BAC ends.
Location/Qualifiers
1..977
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBHe000609"
/tissue_type="immature ear"
/dev_stage="6-8 weeks"
/lab_host="DH10B"
/clone_lib="ZMMBHe"
/note="Vector: TOPOpcr4; Site_1: EcoRI; Site_2: EcoRI"

FEATURES
source
Location/Qualifiers
1..977
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBHe000609"
/tissue_type="immature ear"
/dev_stage="6-8 weeks"
/lab_host="DH10B"
/clone_lib="ZMMBHe"
/note="Vector: TOPOpcr4; Site_1: EcoRI; Site_2: EcoRI"

Query Match 69.3%; Score 20.8; DB 9; Length 977;
Best Local Similarity 91.7%; Pred.No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AACATGATGTCGGGTGTACATC 24
Db 946 AACATGATGTCGGGTGCATC 923

RESULT 7
LOCUS BX507610/c 631 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFP686A14261.r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DKFP686A14261 5', mRNA sequence.
ACCESSION BX507610
VERSION BX507610.1 GI:32040422
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 631)
AUTHORS Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
Wiemann,S.
EST (Poustka,A., Albert,R., Moosmayer,P., Schupp,I.,
Wellenreuther,R., et al.)
Unpublished (2003)
Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by DKFZ (German Cancer Research Center,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project
No g1 sequence available.
This clone (DKFP686A14261) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1..631
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFP686A14261"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiIB;

FEATURES
source
Location/Qualifiers
1..631
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFP686A14261"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiIB;

```



```

TITLE      Expressed genes in Triticum aestivum
JOURNAL    Unpublished (2002)
COMMENT    Contact: Tadasu Shin-i
           Center For Genetic Resource Information
           National Institute of Genetics
           1111 Yata, Mishima, Shizuoka 411-8540, Japan
           Tel: 81-559-81-6856
           Fax: 81-559-81-6855
           Email: tshini@genes.nig.ac.jp.

FEATURES   Location/Qualifiers
   source   1..707
            /organism="Triticum aestivum"
            /mol_type="mRNA"
            /cultivar="Chinese Spring"
            /db_xref="taxon:4565"
            /clone="whyd5n12"
            /tissue_type="spikelet at late flowering"
            /dev_stage="Feekes' scale 6"
            /clone_lib="Y. Ogihara unpublished cDNA library, wh_yd"

ORIGIN
Query Match      68.0%; Score 20.4; DB 4; Length 707;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AACAAATGATGTCGCGGTACATCATGAC 30
    ||||| ||||| ||||| ||||| |||||
Db 222 AACTTGCAATGTCGCGGTACATCATGAC 193

RESULT 11
LOCUS      AG347593               862 bp    DNA    linear    GSS 02-JUN-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-143C18.T7, genomic survey
            sequence.
ACCESSION  AG347593
VERSION    AG347593.1 GI:47920903
KEYWORDS   GSS.
SOURCE     Mus musculus molossinus
   ORGANISM Mus musculus molossinus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1
AUTHORS    Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE      BAC end Sequences of Library MSMg01
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 862)
AUTHORS    Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE      Direct Submission
JOURNAL    Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Saitama-shi, Saitama, Japan
            (E-mail:hattori@psc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
            Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institute, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            Phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : T7
LIBRARY
Vector      : pBACE3.6
R.Site 1    : EcoRI
R.Site 2    : EcoRI.
            Location/Qualifiers
   source   1..862
            /organism="Mus musculus molossinus"
            /mol_type="genomic DNA"
            /sub_species="molossinus"
            /db_xref="taxon:57486"

COMMENT
/clone="MSMg01-143C18.T7"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match      68.0%; Score 20.4; DB 9; Length 862;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AACAAATGATGTCGCGGTACATCATGAC 30
    ||||| ||||| ||||| ||||| |||||
Db 253 AACATAATGTCGCGGTATACATCTGTAC 224

RESULT 12
LOCUS      CL290684/c           960 bp    DNA    linear    GSS 10-FEB-2004
DEFINITION ZMMBB0635A01f ZMMBBB (HindIII) Zea mays genomic clone
            ZMMBB0635A01 5', genomic survey sequence.
ACCESSION  CL290684
VERSION    CL290684.1 GI:42505071
KEYWORDS   GSS.
SOURCE     Zea mays
   ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE  1 (bases 1 to 960)
AUTHORS    Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
            Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
TITLE      Sequencing of the maize genome at PGIR (2003c)
JOURNAL    Unpublished (2003)
COMMENT    Contact: Bharti,A.K.
            Dr.Joachim Messing's lab
            The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
            University
            190 Frelinghuysen Road, Piscataway, NJ 08854, USA
            Tel: 732 445 3801
            Fax: 732 445 5735
            Email: bharti@waksman.rutgers.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 85.
            Location/Qualifiers
   source   1..960
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /cultivar="B73"
            /db_xref="taxon:4577"
            /clone="ZMMBB0635A01"
            /lab_host="E. coli DH10B"
            /clone_lib="ZMMBBB (HindIII)"
            /note="Vector: pCUGI; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      68.0%; Score 20.4; DB 9; Length 960;
Best Local Similarity 80.0%; Pred. No. 1.8e+02;
Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AACAAATGATGTCGCGGTACATCATGAC 30
    ||||| ||||| ||||| ||||| |||||
Db 80 AACAGTCAATGTCGCGGTATACACCGTTAC 51

RESULT 13
LOCUS      BH204000/c           390 bp    DNA    linear    GSS 15-NOV-2002
DEFINITION Sml-54G19.TF Sml Schistosoma mansoni genomic clone Sml-54G19,
            genomic survey sequence.
ACCESSION  BH204000
VERSION    BH204000.1 GI:16378136
KEYWORDS   GSS.

```

```

SOURCE      Schistosoma mansoni
ORGANISM    Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
REFERENCE   1 (bases 1 to 390)
AUTHORS     Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
            Shetty,J., Simpson,A., Malek,J., Koo,H., LoVerde,P.T. and
            El-Sayed,N.W.
TITLE       Use of end sequences from Schistosoma mansoni (Puerto Rico strain)
            Sml BAC library for gene discovery and map construction
JOURNAL     Unpublished (2001)
COMMENT     Other GSSs: Sml-54G19.TR
            Contact: Najib M. El-Sayed
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: nelsayed@tigr.org
            Clones are derived from the Schistosoma mansoni (Puerto Rico) Sml
            BAC library. For clone availability, please contact Dr. Najib
            El-Sayed at TIGR (nelsayed@tigr.org) or Dr. Phillip LoVerde at
            State University of New York, Buffalo, New York, USA
            (loverde@buffalo.edu)
            Seq primer: M13 For
            Class: BAC ends.
FEATURES    source
            1..390
               Location/Qualifiers
               /organism="Schistosoma mansoni"
               /mol_type="genomic DNA"
               /strain="Puerto Rico"
               /db_xref="taxon:6183"
               /clone="Sml-54G19"
               /clone_lib="Sml"
               /note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed
               in the laboratory of Dr. Denis Le Paslier at the Fondation
               Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma
               mansoni agarose embedded DNA was partially digested with
               Hin dIII. High molecular weight fragments were ligated in
               pBelOBAC11 digested with Hin dIII. The average insert size
               is 100 kb. Total clone coverage: approx. 7.95 X the
               haploid genome. Further information can be found in Le
               Paslier et al. (2000) Construction and characterization of
               a Schistosoma mansoni bacterial artificial chromosome
               library. Genomics 65: 87-94."
ORIGIN
Query Match      66.7%; Score 20; DB 8; Length 390;
Best Local Similarity 82.1%; Pred. No. 2.4e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 CAATGTATGTCGGGTGACATCTATGAC 30
    ||||| ||||| ||||| ||||| |||||
Db 387 CAATGTTGTCGGGTGATGATGAC 360

RESULT 14
LOCUS      CL412828
DEFINITION RCI44_431D15.r RCI44-44 Sus scrofa genomic clone RCI44_431D15,
            genomic survey sequence.
ACCESSION  CL412828
VERSION     CL412828.1 GI:51446888
KEYWORDS   GSS.
SOURCE     Sus scrofa (pig)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 538)
AUTHORS     Rogatcheva,M.B., Meyers,S., He,W., Larkin,D.M., Marron,B.M.,
            Baeverf.J.E. and Schook,L.B.
TITLE       P959-BACing the Human Genome: Constructing a Porcine Physical Map
            Through Comparative Genomics
JOURNAL     Unpublished (2004)

Other_GSSs: RCI44_431D15.f
Contact: Lawrence B. Schook
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 265 5326
Fax: 217 244 5617
Email: schook@uiuc.edu
Clones are derived from the porcine BAC library RCI44-44
(http://www.bacpac.chori.org/porcine242.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@chori.org).
Clones may be purchased from BACPAC Resources
(http://BACPACresources.chori.org). This work was undertaken as part
of the International Swine Genome Sequencing Consortium by
University of Illinois at Urbana Champaign, USA with funds provided
by grant No. AG2002-34480-11828 from USDA-CSREES and
AG2001-35205-09965 from USDA/NRI (Livestock Genome Sequencing
Initiative)
Plate: 431 row: D column: 15
Seq primer: SP6
Class: BAC ends.
FEATURES    Location/Qualifiers
            1..538
               /organism="Sus scrofa"
               /mol_type="genomic DNA"
               /strain="four pigs (breed: 37.5% Yorks Landrace and 25%
               Meishan)"
               /db_xref="taxon:9823"
               /clone="RCI44_431D15"
               /sex="male"
               /cell_type="blood"
               /clone_lib="RCI44-44"
               /notes="Vector: pTARBAC2; Site 1: EcoRI; Site 2: EcoRI;
               porcine male BAC library produced by Pieter de Jong"
ORIGIN
Query Match      66.7%; Score 20; DB 9; Length 538;
Best Local Similarity 82.1%; Pred. No. 2.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AACATGTATGTCGGGTGATCATCTATG 28
    ||||| ||||| ||||| ||||| |||||
Db 2 AACATGTATGTCGGGTGATCATCTATG 29

RESULT 15
LOCUS      AQ879634/c
DEFINITION HS_4816_A2_H10_T7A CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=4816 Col=20 Row=0, genomic survey
            sequence.
ACCESSION  AQ879634
VERSION     AQ879634.1 GI:6311101
KEYWORDS   GSS.
SOURCE     Homo sapiens (human)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 542)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,K., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE    93380589
PUBMED     10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887

```

Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 4816 row: O column: 20
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 542.

FEATURES source

Location/Qualifiers
 1. .542
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone_lib="plate:4816 Col=20 Row=O"
 /sex="male"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

ORIGIN

Query Match 66.7%; Score 20; DB 8; Length 542;
 Best Local Similarity 82.1%; Pred. No. 2.5e+02;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACATGTATGTCGCGTGTACATCTATGA 29
 Db 157 ACAGTGAAGTAGGTGTACACCTATGA 130

RESULT 16

AQ525705/c
 LOCUS
 DEFINITION HS 5239_B1_C02_T7A RPCI-11 Human Male BAC Library GSS 11-MAY-1999
 genomic_clone Plate=815 Col=3 Row=F, genomic survey sequence.

ACCESSION AQ525705
 VERSION AQ525705.1 GI:4773025
 KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 605)

REFERENCE
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL
 MEDLINE
 PUBMED
 COMMENT Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887

Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 815 row: F column: 3
 Seq primer: T7

Class: BAC ends
 High quality sequence stop: 605.

FEATURES source

Location/Qualifiers
 1. .605
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone="plate:815 Col=3 Row=F"
 /sex="male"

ORIGIN

Query Match 66.7%; Score 20; DB 8; Length 605;
 Best Local Similarity 82.1%; Pred. No. 2.6e+02;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 CAATGTATGTCGCGTGTACATCTATGAC 30
 Db 535 CAATGTATGTGAGTGTGATGTATGAC 508

RESULT 17

AQ856387
 LOCUS
 DEFINITION

AQ856387 773 bp DNA linear GSS 03-NOV-1999
 nbeb0003A23r CUGI Rice BAC Library (EcoRI) Oryza sativa (japonica cultivar-group) genomic clone nbeb0003A23r, genomic survey sequence.

ACCESSION AQ856387
 VERSION AQ856387.1 GI:6206844
 KEYWORDS GSS.

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 773)

AUTHORS Wing,R.A. and Dean,R.A.

TITLE A BAC End Sequencing Framework to Sequence the Rice Genome

JOURNAL Unpublished (1998)

COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: GGAAACAGCTATGACCATG

Class: BAC ends

High quality sequence start: 8

High quality sequence stop: 311.

FEATURES source

Location/Qualifiers
 1. .773
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="nbeb0003A23r"
 /tissue_type="Leaf"
 /lab_host="E. coli DH10B"
 /clone_lib="CUGI Rice BAC Library (EcoRI)"
 /note="Vector: pBACIndigo; Site 1: EcoRI; Site 2: EcoRI;
 Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %.

ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 1330)
 Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and Ecker, J.
 BAC End Sequences at ATGC
 Unpublished (1997)
 Other GSSs: F26A16-T7.1, F26A16-T7.2, F26A16-Sp6, F26A16-Sp6.1
 Contact: Ecker J.
 Arabidopsis Thaliana Genome Center
 University of Pennsylvania
 Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104
 Tel: 215-898-9384
 Fax: 215-898-8780
 Email: jecker@genome.bio.upenn.edu
 Seq primer: T7
 Class: BAC ends
 High quality sequence start: 92
 High quality sequence stop: 678.
 Location/Qualifiers
 1. .1330
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Columbia"
 /db_xref="taxon:3702"
 /clone="F26A16"
 /sex="hermaphrodite"
 /clone_lib="IGF"
 /note="Vector: BelOBACII; Site_1: EcoRI; Site_2: EcoRI; Produced by Thomas Altmann"

ORIGIN

Query Match 66.7%; Score 20; DB 8; Length 1330;
 Best Local Similarity 82.1%; Pred. No. 2.9e+02;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACAATGATATGTCGGTGTACATCTATGA 29
 ||||| ||||| ||||| ||||| |||||
 Db 136 ACAATATATGACTGTGTAAACCTATGA 109

RESULT 21
 BZ091289
 LOCUS
 DEFINITION CH230-213G14-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
 CH230-213G14, genomic survey sequence.
 ACCESSION BZ091289
 VERSION BZ091289.1 GI:23728348
 KEYWORDS GSS.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 266)
 Zhao, S., Shetty, J., Shatsman, S., Tsengaye, G., Geer, K., Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.
 Rat BAC End Sequences from Library CHORI-230 EcoRI segment
 Unpublished (1999)
 Other GSSs: CH230-213G14.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (http://www.chori.org/bacpac/rat230.htm). For BAC library

ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 1330)
 Feng, J., Dewar, K., Buehler, E., Kim, C., Li, Y., Shinn, P., Sun, H. and Ecker, J.
 BAC End Sequences at ATGC
 Unpublished (1997)
 Other GSSs: F26A16-T7.1, F26A16-T7.2, F26A16-Sp6, F26A16-Sp6.1
 Contact: Ecker J.
 Arabidopsis Thaliana Genome Center
 University of Pennsylvania
 Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104
 Tel: 215-898-9384
 Fax: 215-898-8780
 Email: jecker@genome.bio.upenn.edu
 Seq primer: T7
 Class: BAC ends
 High quality sequence start: 92
 High quality sequence stop: 678.
 Location/Qualifiers
 1. .1330
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Columbia"
 /db_xref="taxon:3702"
 /clone="F26A16"
 /sex="hermaphrodite"
 /clone_lib="IGF"
 /note="Vector: BelOBACII; Site_1: EcoRI; Site_2: EcoRI; Produced by Thomas Altmann"

ORIGIN

Query Match 66.7%; Score 20; DB 8; Length 1330;
 Best Local Similarity 82.1%; Pred. No. 2.9e+02;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 ACAATGATATGTCGGTGTACATCTATGA 29
 ||||| ||||| ||||| ||||| |||||
 Db 136 ACAATATATGACTGTGTAAACCTATGA 109

RESULT 21
 BZ091289
 LOCUS
 DEFINITION CH230-213G14-TV CHORI-230 Segment 1 Rattus norvegicus genomic clone
 CH230-213G14, genomic survey sequence.
 ACCESSION BZ091289
 VERSION BZ091289.1 GI:23728348
 KEYWORDS GSS.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 266)
 Zhao, S., Shetty, J., Shatsman, S., Tsengaye, G., Geer, K., Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.
 Rat BAC End Sequences from Library CHORI-230 EcoRI segment
 Unpublished (1999)
 Other GSSs: CH230-213G14.TJ
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (http://www.chori.org/bacpac/rat230.htm). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/or ering_information.htm). BAC end
 plate: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
 Plate: 213 row: G column: 14
 Seq primer: T7
 Class: BAC ends.
 Location/Qualifiers
 1. .266
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SnHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-213G14"
 /sex="Female"
 /cell_type="Brain"
 /clone_lib="CHORI-230 Segment 1"
 /notes="Vector: PTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
 CHORI-230 Rat (BN/SnHsd/MCW) BAC library produced by
 Pieter de Jong"

ORIGIN

Query Match 65.3%; Score 19.6; DB 8; Length 266;
 Best Local Similarity 84.6%; Pred. No. 3.4e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 ATGTATGTCGGTGTACATCTATGAC 30
 ||||| ||||| ||||| ||||| |||||
 Db 148 ATGTGTGTCGGTGTACATATATGTC 173

RESULT 22
 AA751822
 LOCUS
 DEFINITION 96AS0663 Rice Immature Seed Lambda ZAPII cDNA Library Oryza sativa
 (indica cultivar-group) cDNA clone 96AS0663, mRNA sequence.
 AA751822
 ACCESSION AA751822
 VERSION AA751822.1 GI:2798528
 KEYWORDS EST.
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.
 1 (bases 1 to 380)
 Nahm, B.H., Kim, J.K., Cheong, J.J., Kim, S.I., Hahn, T.R., Moon, E.P.,
 Kim, W.T., Kim, W.Y., Yang, M.S., Park, R.D., Sohn, U.I., Kang, K.Y.,
 Lee, M.C. and Eun, M.Y.
 Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
 Unpublished (1998)
 Contact: Eun M.Y.
 Department of Cytogenetics
 National Inst. of Agri. Sci. and Tech, RDA
 Suwon, Kyunggido, Korea
 Tel: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeun@sun20.asti.re.kr
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
 University, Yongin, Korea 449-728 bhna@biolserver.myongji.ac.kr
 Seq primer: M13 Reverse Primer.
 Location/Qualifiers
 1. .380
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Milyang 23"
 /db_xref="taxon:39946"
 /clone="96AS0663"
 /tissue_type="Immature Seed"
 /dev_stage="5 days after pollination"
 /lab_host="E. coli SOLR"
 /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
 /notes="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
 XhoI; Directional cDNA library inserted into lambda ZAPII

FEATURES
 source
 1. .380
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Milyang 23"
 /db_xref="taxon:39946"
 /clone="96AS0663"
 /tissue_type="Immature Seed"
 /dev_stage="5 days after pollination"
 /lab_host="E. coli SOLR"
 /clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
 /notes="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
 XhoI; Directional cDNA library inserted into lambda ZAPII

vector at 5' end with EcoRI and 3' end with Xho I site."

ORIGIN
Query Match 65.3%; Score 19.6; DB 1; Length 380;
Best Local Similarity 78.6%; Pred. No. 3.6e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGATCATCTATG 28
|||||
Db 327 AAAAATGCATCTNANGTGATCATCTATG 354
|||||

RESULT 23
BH121074
LOCUS
DEFINITION BHI21074 385 bp DNA linear GSS 19-JUL-2001
genomic survey sequence.
ACCESSION BHI21074
VERSION BHI21074.1 GI:14964586
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 385)
AUTHORS Zhao S., Nierman, W., Malek, J., Shatsman, S., Akimret, B., Levins, M., Teagaye, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregorgis, E., Russell, D., de Jong, P. and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-24
JOURNAL Unpublished (1999)
COMMENT Other GSSs: RPCI-24-230E9.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.choi.org/bacpac/orderingframe.html>). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 230 row: E column: 9
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..385
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-230E9"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/note="Vector: pTARBAC1; Site 1: BamH1; Site 2: BamH1; RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamH1 sites using MboI partially digested male C57BL/6J DNA."

ORIGIN
Query Match 65.3%; Score 19.6; DB 8; Length 385;
Best Local Similarity 84.6%; Pred. No. 3.6e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ACAATGTATGTCGGGTGATCATCTAT 27
|||||
Db 166 ACAGTGTATGTCCTGCTACATTTAT 191
|||||

RESULT 24

CF889471
LOCUS
DEFINITION TCTR-272 TcTR Trypanosoma cruzi cDNA clone 02e8 5', mRNA sequence.
ACCESSION CF889471
VERSION CF889471.1 GI:44919354
KEYWORDS EST.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi

REFERENCE 1 (bases 1 to 396)
AUTHORS Agüero, F., Ben Abdallah, K., Tekiel, V., Sanchez, D.O. and Gonzalez, A.
TITLE Generation and analysis of expressed sequence tags from Trypanosoma cruzi trypanastigote and amastigote cDNA libraries
JOURNAL Mol. Biochem. Parasitol. 136 (2), 221-225 (2004)
COMMENT Contact: Sanchez DO
Genomics and Bioinformatics
Instituto de Investigaciones Biotecnológicas
Av. Gral Paz S/N, INTI, Edificio 24, B 1650 KNA, San Martín, Buenos Aires, Argentina
Tel: (54-11) 4580/7255/7
Fax: (54-11) 4752-9639
Email: dsanchez@ib.unsam.edu.ar

Sequences were basecalled with phred and vector was masked with crossmatch (see <http://www.phrap.org>). Sequences were then trimmed from both ends to remove low quality bases and masked vector.
Plate: 02 row: e column: 8
Seq primer: T7.
Location/Qualifiers
1..396
/organism="Trypanosoma cruzi"
/mol_type="mRNA"
/db_xref="taxon:5693"
/clone="02e8"
/dev_stage="trypanastigote"
/clone_lib="TcTR"
/note="Vector: pBluescript KS+; Site_1: EcoRI; Site_2: NotI; PolyA(+) RNA was prepared using the QuickPrep Micro RNA Purification kit (Amersham Pharmacia Biotech). Synthesis of the cDNA first strand was done with Superscript II reverse transcriptase (Gibco BRL) and an oligo-dT-Not primer (5'-CTGCGCGCGCT(18)-3'). Synthesis of the cDNA second strand was performed using the Klenow fragment of DNA polymerase with a spliced leader (SL) primer (5'-GATACAGTTCTGTA-3'). After methylation with EcoRI methylase, phosphorylated EcoRI linkers (5'-ACGGAATTCGT-3') were ligated to the cDNA. The resulting cDNA mixture was then digested with NotI and EcoRI restriction enzymes, subjected to size fractionation on SizeSep400 Spun Columns (Pharmacia) to select fragments longer than 400 bp, and cloned into the dephosphorylated NotI/EcoRI sites of the vector."

FEATURES
source Location/Qualifiers
1..396
/organism="Trypanosoma cruzi"
/mol_type="mRNA"
/db_xref="taxon:5693"
/clone="02e8"
/dev_stage="trypanastigote"
/clone_lib="TcTR"
/note="Vector: pBluescript KS+; Site_1: EcoRI; Site_2: NotI; PolyA(+) RNA was prepared using the QuickPrep Micro RNA Purification kit (Amersham Pharmacia Biotech). Synthesis of the cDNA first strand was done with Superscript II reverse transcriptase (Gibco BRL) and an oligo-dT-Not primer (5'-CTGCGCGCGCT(18)-3'). Synthesis of the cDNA second strand was performed using the Klenow fragment of DNA polymerase with a spliced leader (SL) primer (5'-GATACAGTTCTGTA-3'). After methylation with EcoRI methylase, phosphorylated EcoRI linkers (5'-ACGGAATTCGT-3') were ligated to the cDNA. The resulting cDNA mixture was then digested with NotI and EcoRI restriction enzymes, subjected to size fractionation on SizeSep400 Spun Columns (Pharmacia) to select fragments longer than 400 bp, and cloned into the dephosphorylated NotI/EcoRI sites of the vector."

ORIGIN
Query Match 65.3%; Score 19.6; DB 7; Length 396;
Best Local Similarity 84.6%; Pred. No. 3.7e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 CAATGTATGTCGGGTGATCATCTATG 28
|||||
Db 172 CAATGTGTGTCCTCATGTATCTATG 197
|||||

RESULT 25
AI562331
LOCUS
DEFINITION TENS2326 T. cruzi epimastigote normalized cDNA Library Trypanosoma cruzi cDNA clone 2326 5', mRNA sequence.
ACCESSION AI562331
VERSION AI562331.1 GI:4513676
KEYWORDS EST.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi

Sequences were basecalled with phred and vector was masked with crossmatch (see <http://www.phrap.org>). Sequences were then trimmed from both ends to remove low quality bases and masked vector.
Plate: 02 row: e column: 8
Seq primer: T7.
Location/Qualifiers
1..396
/organism="Trypanosoma cruzi"
/mol_type="mRNA"
/db_xref="taxon:5693"
/clone="02e8"
/dev_stage="trypanastigote"
/clone_lib="TcTR"
/note="Vector: pBluescript KS+; Site_1: EcoRI; Site_2: NotI; PolyA(+) RNA was prepared using the QuickPrep Micro RNA Purification kit (Amersham Pharmacia Biotech). Synthesis of the cDNA first strand was done with Superscript II reverse transcriptase (Gibco BRL) and an oligo-dT-Not primer (5'-CTGCGCGCGCT(18)-3'). Synthesis of the cDNA second strand was performed using the Klenow fragment of DNA polymerase with a spliced leader (SL) primer (5'-GATACAGTTCTGTA-3'). After methylation with EcoRI methylase, phosphorylated EcoRI linkers (5'-ACGGAATTCGT-3') were ligated to the cDNA. The resulting cDNA mixture was then digested with NotI and EcoRI restriction enzymes, subjected to size fractionation on SizeSep400 Spun Columns (Pharmacia) to select fragments longer than 400 bp, and cloned into the dephosphorylated NotI/EcoRI sites of the vector."

ORIGIN
Query Match 65.3%; Score 19.6; DB 7; Length 396;
Best Local Similarity 84.6%; Pred. No. 3.7e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 CAATGTATGTCGGGTGATCATCTATG 28
|||||
Db 172 CAATGTGTGTCCTCATGTATCTATG 197
|||||

RESULT 25
AI562331
LOCUS
DEFINITION TENS2326 T. cruzi epimastigote normalized cDNA Library Trypanosoma cruzi cDNA clone 2326 5', mRNA sequence.
ACCESSION AI562331
VERSION AI562331.1 GI:4513676
KEYWORDS EST.
SOURCE Trypanosoma cruzi
ORGANISM Trypanosoma cruzi

ORIGIN
Query Match 65.3%; Score 19.6; DB 8; Length 385;
Best Local Similarity 84.6%; Pred. No. 3.6e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 ACAATGTATGTCGGGTGATCATCTAT 27
|||||
Db 166 ACAGTGTATGTCCTGCTACATTTAT 191
|||||

RESULT 24

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma; Schizotrypanum.
1 (bases 1 to 475)
Verdun,R.E., Di Paolo,N.C., Urmenyi,T.P., Rondinelli,E., Frasci,A.C.C. and Sanchez,D.O.
Gene discovery through expressed sequence tag sequencing in trypanosoma cruzi
Infect. Immun. 66 (11), 5393-5398 (1998)
99003155
Contact: Sanchez D.O.
Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral San Martin)
Av. Gral Paz entre Albarcellos y Constituyentes, INTI edificio 24 Cp(1650) San Martin, Prov. de BS AS, Argentina
Tel: (54-1)752-9639 or (54-1)752-0021
Fax: (54-1)752-0021 or (54-1)752-9639
Email: dsanchez@inti.gov.ar
Seq primer: T7.
Location/Qualifiers
1..475
/organism="Trypanosoma cruzi"
/mol_type="mRNA"
/strain="Cl-Brenner"
/db_xref="taxon:5693"
/clone="2326"
/cell_type="epimastigote"
/clone_lib="T. cruzi epimastigote normalized cDNA Library"
/note="cDNA library constructed with oligo dT primed epimastigote mRNA and cloned in pT7b18D phagemid with modified polylinker (PHARMACIA)"

Query Match 65.3%; Score 19.6; DB 1; Length 475;
Best Local Similarity 84.6%; Pred. No. 3.8e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CAATCTATGTCGGGTGACATCTATG 28
|||||
Db 133 CAATGTTTGCCCAAGTACATCTATG 158
|||||

RESULT 26
BZ251753/c
LOCUS
DEFINITION
BZ251753 561 bp DNA linear GSS 12-OCT-2002
CH230-306K2.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-306K2, genomic survey sequence.
BZ251753
BZ251753.1 GI:23913040
GSS.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 561)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shwartsbeyn,A., Gbregorgeis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment Unpublished (1999)
Other GSSs: CH230-306K2.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.choi.org). Clones may be purchased from BACPAC Resources

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Qy 1 AACAACTATGTCGCGGTACATCTA 26
 |||||
 Db 92 AACAACTGTTGTCGCGGTGGACCTA 67
 |||||
 RESULT 28
 CL716190
 LOCUS
 DEFINITION CL716190 700 bp DNA linear GSS 26-JUL-2004
 OR_BBa0042K21.r OR_BBa Oryza rufipogon genomic clone OR_BBa0042K21
 3'-genomic survey' sequence.
 ACCESSION CL716190
 VERSION CL716190.1 GI:50603228
 KEYWORDS GSS.
 SOURCE
 ORGANISM
 Oryza rufipogon
 Oryza rufipogon
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
 Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
 OMAP Project
 Unpublished (2004)
 CONTACT: Rod A. Wing
 Arizona Genomics Institute
 University of Arizona
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA
 Tel: 520 626 9595
 Fax: 520 621 1259
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: TAA TAC GAC TCA CTA TAG GG
 BACKWARD: CAC TCA TTA GGC ACC CCA
 Insert Length: 161 Std Error: 0.00
 Plate: 0042 row: K Column: 21
 Seq primer: CAC TCA TTA GGC ACC CCA
 Class: BAC ends.
 Location/Qualifiers
 1..700
 /organism="Oryza rufipogon"
 /mol_type="genomic DNA"
 /db_xref="taxon:4529"
 /clone="OR_BBa0042K21"
 /tissue_type="young leaves"
 /lab_host="DH10B-T1 phage resistant"
 /clone_lib="OR_BBa"
 /note="vector: pGIBAC1; Site_1: HindIII; Site_2: HindIII"

Qy 2 ACAATGATGTCGCGGTACATCTAT 27
 |||||
 Db 445 AAAATTTATCTCGGTGTACATCTAT 470
 |||||
 RESULT 29
 CL507353/c
 LOCUS
 DEFINITION SAIL_780_F03.v1 SAIL Collection Arabidopsis thaliana genomic clone
 SAIL_780_F03.v1, genomic survey sequence.
 ACCESSION CL507353
 VERSION CL507353.1 GI:46004673
 KEYWORDS GSS.
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 1 (bases 1 to 905)

Qy 1 AACAACTATGTCGCGGTACATCTA 26
 |||||
 Db 92 AACAACTGTTGTCGCGGTGGACCTA 67
 |||||
 RESULT 28
 CL716190
 LOCUS
 DEFINITION CL716190 700 bp DNA linear GSS 26-JUL-2004
 OR_BBa0042K21.r OR_BBa Oryza rufipogon genomic clone OR_BBa0042K21
 3'-genomic survey' sequence.
 ACCESSION CL716190
 VERSION CL716190.1 GI:50603228
 KEYWORDS GSS.
 SOURCE
 ORGANISM
 Oryza rufipogon
 Oryza rufipogon
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
 Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
 OMAP Project
 Unpublished (2004)
 CONTACT: Rod A. Wing
 Arizona Genomics Institute
 University of Arizona
 Forbes Building Room 303, Tucson, AZ 85721-0036, USA
 Tel: 520 626 9595
 Fax: 520 621 1259
 Email: http://genome.arizona.edu
 PCR Primers
 FORWARD: TAA TAC GAC TCA CTA TAG GG
 BACKWARD: CAC TCA TTA GGC ACC CCA
 Insert Length: 161 Std Error: 0.00
 Plate: 0042 row: K Column: 21
 Seq primer: CAC TCA TTA GGC ACC CCA
 Class: BAC ends.
 Location/Qualifiers
 1..700
 /organism="Oryza rufipogon"
 /mol_type="genomic DNA"
 /db_xref="taxon:4529"
 /clone="OR_BBa0042K21"
 /tissue_type="young leaves"
 /lab_host="DH10B-T1 phage resistant"
 /clone_lib="OR_BBa"
 /note="vector: pGIBAC1; Site_1: HindIII; Site_2: HindIII"

Qy 2 ACAATGATGTCGCGGTACATCTAT 27
 |||||
 Db 445 AAAATTTATCTCGGTGTACATCTAT 470
 |||||
 RESULT 29
 CL507353/c
 LOCUS
 DEFINITION SAIL_780_F03.v1 SAIL Collection Arabidopsis thaliana genomic clone
 SAIL_780_F03.v1, genomic survey sequence.
 ACCESSION CL507353
 VERSION CL507353.1 GI:46004673
 KEYWORDS GSS.
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 1 (bases 1 to 905)

AUTHORS Sessions,A., Burke,B., Presting,G., Aux,G., McElver,J., Patton,D., Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D., Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmery,B., Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
 A high-throughput Arabidopsis reverse genetics system
 Plant Cell 14 (12), 2985-2994 (2002)
 22356987
 12468722
 COMMENT
 CONTACT: Sessions A
 Applied Trait Genetics
 Syngenta Biotechnology Inc.
 3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
 Email: allen.sessions@syngenta.com
 ABRC Stock Number CS934911; T-DNA left border flanking sequences of Syngenta Arabidopsis Insertion Library (SAIL) lines are available through the Arabidopsis Biological Resource Center (ABRC).
 Sequences represent a pool of amplified genomic regions and not single contiguous sequences.
 Class: TDNA tagged.
 Location/Qualifiers
 1..905
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /ecotype="Columbia"
 /db_xref="taxon:3702"
 /clone="SAIL_780_F03.v1"
 /clone_lib="SAIL Collection"
 /note="T-DNA left border sequences were isolated using a modified TAIL-PCR strategy"

ORIGIN
 Query Match 65.3%; Score 19.6; DB 9; Length 905;
 Best Local Similarity 84.6%; Pred. No. 4.2e+02;
 Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 5 ATGTATGTCGCGGTACATCTATGAC 30
 |||||
 Db 131 AAGTACGTCGCGGTACATCTATGAC 106
 |||||
 RESULT 30
 BF249864
 LOCUS
 DEFINITION BF249864 360 bp mRNA linear EST 10-MAY-2001
 pa87508.y1 Hawdon Ancylostoma caninum L3 Ancylostoma caninum cDNA
 5' similar to SW:UN97_CABEL P50464 LIM PROTEIN UNC-97. [1]; mRNA
 sequence.
 BF249864
 BF249864.1 GI:11180047
 EST.
 Ancylostoma caninum (dog hookworm)
 Ancylostoma caninum
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
 1 (bases 1 to 360)
 McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Teagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterson,R. and Wilson,R.
 The Washington Univ. Nematode EST Project, 1999
 Unpublished (1999)
 CONTACT: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: eet@wustl.edu
 The library was constructed by Dr. John Hawdon of Yale University
 (John.hawdon@yale.edu). DNA Sequencing by: Washington University
 Genome Sequencing Center, St. Louis.

TITLE
 JOURNAL
 COMMENT

FEATURES
source

Location/Qualifiers
1. .360
/organism="Ancylostoma caninum"
/mol_type="mRNA"
/strain="Shanghai strain"
/db_xref="taxon:29170"
/sex="female and male"
/dev_stage="3rd stage larva (L3)"
/lab_host="E. coli, XL-1 Blue MRF, (Stratagene)"
/clone_lib="Hawdon Ancylostoma caninum L3"
/note="Vector: Lambda Uni-ZAP XR Vector (Stratagene);
Site 1: EcoRI (5'); Site 2: XhoI (3'); Oligo (dT) primed
library from L3 worms isolated from dog. cDNA was
constructed and cloned unidirectionally into the vector
from the EcoRI to the XhoI site. The library went
through one round of amplification."

ORIGIN

Query Match 64.7%; Score 19.4; DB 2; Length 360;
Best Local Similarity 79.3%; Pred. No. 4.4e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 AACAAATGATGTCGGGTGATCATCTATGA 29
Db 298 AACAGATATATCCGGAGCTCATCTATGA 326

Search completed: March 18, 2005, 10:18:59
Job time : 3121 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2005, 08:42:23 ; Search time 128 Seconds
(without alignments)

383.502 Million cell updates/sec

Title: US-09-806-197-24

Perfect score: 30

Sequence: 1 aacaatgtatgtccggtgtacatctatgac 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.6	78.7	2042	2	US-08-911-434A-2
2	19	63.3	198632	4	US-08-949-016-12781
3	19	63.3	198637	4	US-08-949-016-17393
c 4	18.6	62.0	601	4	US-08-949-016-103581
5	18.6	62.0	2152	4	US-08-949-016-13851
c 6	18.6	62.0	256287	4	US-08-949-016-14608
c 7	18.4	61.3	40512	4	US-08-949-016-16612
c 8	18	60.0	6439	4	US-08-902-540-813
9	17.8	59.3	438	3	US-08-651-155B-134
10	17.8	59.3	438	4	US-08-194-036B-134
11	17.8	59.3	908	3	US-08-457-046B-15
12	17.8	59.3	908	4	US-08-866-570B-15
13	17.8	59.3	1686	4	US-08-902-540-9105
14	17.8	59.3	10318	4	US-08-902-540-973
15	17.8	59.3	19698	4	US-08-949-016-14079
c 16	17.8	59.3	40742	4	US-08-949-016-11751
c 17	17.8	59.3	40742	4	US-08-949-016-13097
c 18	17.8	59.3	70088	4	US-08-949-016-16845
c 19	17.8	59.3	123463	4	US-08-949-016-17078
20	17.8	59.3	235452	4	US-08-949-016-13675
21	17.6	58.7	601	4	US-08-949-016-32764
22	17.6	58.7	601	4	US-08-949-016-32823
23	17.6	58.7	601	4	US-08-949-016-50681
24	17.6	58.7	601	4	US-08-949-016-50740
25	17.6	58.7	27702	4	US-08-949-016-11795
c 26	17.6	58.7	77851	4	US-08-949-016-12508
c 27	17.6	58.7	77867	4	US-08-949-016-13211
1	23.6	78.7	2042	2	US-08-911-434A-2
2	19	63.3	198632	4	US-08-949-016-12781
3	19	63.3	198637	4	US-08-949-016-17393
c 4	18.6	62.0	601	4	US-08-949-016-103581
5	18.6	62.0	2152	4	US-08-949-016-13851
c 6	18.6	62.0	256287	4	US-08-949-016-14608
c 7	18.4	61.3	40512	4	US-08-949-016-16612
c 8	18	60.0	6439	4	US-08-902-540-813
9	17.8	59.3	438	3	US-08-651-155B-134
10	17.8	59.3	438	4	US-08-194-036B-134
11	17.8	59.3	908	3	US-08-457-046B-15
12	17.8	59.3	908	4	US-08-866-570B-15
13	17.8	59.3	1686	4	US-08-902-540-9105
14	17.8	59.3	10318	4	US-08-902-540-973
15	17.8	59.3	19698	4	US-08-949-016-14079
c 16	17.8	59.3	40742	4	US-08-949-016-11751
c 17	17.8	59.3	40742	4	US-08-949-016-13097
c 18	17.8	59.3	70088	4	US-08-949-016-16845
c 19	17.8	59.3	123463	4	US-08-949-016-17078
20	17.8	59.3	235452	4	US-08-949-016-13675
21	17.6	58.7	601	4	US-08-949-016-32764
22	17.6	58.7	601	4	US-08-949-016-32823
23	17.6	58.7	601	4	US-08-949-016-50681
24	17.6	58.7	601	4	US-08-949-016-50740
25	17.6	58.7	27702	4	US-08-949-016-11795
c 26	17.6	58.7	77851	4	US-08-949-016-12508
c 27	17.6	58.7	77867	4	US-08-949-016-13211

Sequence 13212, A
Sequence 12509, A
Sequence 31, Appl
Sequence 34, Appl
Sequence 7260, Ap
Sequence 15, Appl
Sequence 916, App
Sequence 8, Appl
Sequence 28, Appl
Sequence 14, Appl
Sequence 777, App
Sequence 777, App
Sequence 777, App
Sequence 777, App
Sequence 777, App
Sequence 17, Appl
Sequence 182, App
Sequence 71655, A
Sequence 71656, A
Sequence 71728, A
Sequence 71729, A
Sequence 71801, A
Sequence 71802, A
Sequence 71874, A
Sequence 71875, A
Sequence 71947, A
Sequence 71948, A
Sequence 72020, A
Sequence 72021, A
Sequence 72093, A
Sequence 72094, A
Sequence 72166, A
Sequence 72167, A
Sequence 72239, A
Sequence 72240, A
Sequence 72312, A
Sequence 72313, A
Sequence 2, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 1, Appl
Sequence 7, Appl
Sequence 13849, A
Sequence 13850, A
Sequence 13851, A
Sequence 13852, A
Sequence 13853, A
Sequence 13854, A
Sequence 13855, A
Sequence 13856, A
Sequence 13857, A
Sequence 13858, A
Sequence 17057, A
Sequence 31, Appl
Sequence 34, Appl
Sequence 205744, A
Sequence 205745, A
Sequence 205746, A
Sequence 12913, A
Sequence 4, Appl
Sequence 16402, A
Sequence 13683, A
Sequence 13194, A
Sequence 13481, A
Sequence 12705, A
Sequence 17550, A
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl

```
101 16.8 56.0 509 4 US-09-949-016-135861
102 16.8 56.0 509 4 US-09-949-016-135978
103 16.8 56.0 509 4 US-09-949-016-136095
104 16.8 56.0 601 4 US-09-949-016-136095
105 16.8 56.0 601 4 US-09-949-016-136095
106 16.8 56.0 601 4 US-09-949-016-136095
107 16.8 56.0 601 4 US-09-949-016-136095
108 16.8 56.0 601 4 US-09-949-016-136095
109 16.8 56.0 601 4 US-09-949-016-136095
110 16.8 56.0 601 4 US-09-949-016-136095
111 16.8 56.0 1529 3 US-08-821-278A-1
112 16.8 56.0 1609 4 US-09-449-632-1
113 16.8 56.0 2166 4 US-09-107-532A-3235
114 16.8 56.0 4819 4 US-09-774-528-72
115 16.8 56.0 5707 2 US-08-472-809B-8
116 16.8 56.0 6345 2 US-08-472-809B-7
117 16.8 56.0 83218 4 US-09-949-016-14489
118 16.8 56.0 105679 4 US-09-949-016-12360
119 16.8 56.0 119801 4 US-09-949-016-13453
120 16.8 56.0 126468 4 US-09-949-016-14418
121 16.8 56.0 131724 4 US-09-949-016-12893
122 16.8 56.0 186734 4 US-09-949-016-14870
123 16.8 56.0 193555 4 US-09-949-016-15553
124 16.8 56.0 193555 4 US-09-949-016-15554
125 16.8 56.0 193555 4 US-09-949-016-12350
126 16.8 56.0 193689 4 US-09-949-016-13088
127 16.8 56.0 213456 4 US-09-820-007-3
128 16.8 56.0 236341 4 US-09-949-016-13978
129 16.8 56.0 323820 4 US-09-949-016-14139
130 16.8 56.0 451924 4 US-09-949-016-12896
131 16.8 56.0 451925 4 US-09-949-016-17305
132 16.8 56.0 640681 4 US-09-790-988-1
133 16.6 55.3 601 4 US-09-949-016-55493
134 16.6 55.3 601 4 US-09-949-016-137549
135 16.6 55.3 601 4 US-09-949-016-137550
136 16.6 55.3 91665 4 US-09-949-016-12234
137 16.6 55.3 96739 4 US-09-949-016-15606
138 16.6 55.3 113379 4 US-09-949-016-17561
139 16.6 55.3 113379 4 US-09-949-016-17562
140 16.6 55.3 285478 4 US-09-949-016-13362
141 16.6 55.3 422592 4 US-09-949-016-14182
142 16.6 55.3 767677 4 US-09-949-016-12147
143 16.6 55.3 767677 4 US-09-949-016-17361
144 16.6 54.7 148 4 US-09-513-999C-30057
145 16.4 54.7 484 3 US-08-945-983-3
146 16.4 54.7 484 3 US-08-945-983-5
147 16.4 54.7 600 6 5198345-4
148 16.4 54.7 600 6 5198345-4
149 16.4 54.7 601 4 US-09-949-016-19071
150 16.4 54.7 601 4 US-09-949-016-36885

ALIGNMENTS

RESULT 1
US-08-911-434A-2
; Sequence 2, Application US/08911434A
; Patent No. 5959176
; GENERAL INFORMATION:
; APPLICANT: TORIKAI, Satomi
; APPLICANT: OEDA, Kenji
; TITLE OF INVENTION: PLANT PROMOTER AND UTILIZATION THEREOF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VIRGINIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 22040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

Sequence 135861,
Sequence 135978,
Sequence 136095,
Sequence 136095,
Sequence 136095,
Sequence 136095,
Sequence 136095,
Sequence 136095,
Sequence 136095,
Sequence 136095,
Sequence 1, Appli
Sequence 1, Appli
Sequence 3235, Ap
Sequence 72, Appli
Sequence 8, Appli
Sequence 7, Appli
Sequence 14489, A
Sequence 12360, A
Sequence 13453, A
Sequence 14418, A
Sequence 12893, A
Sequence 14870, A
Sequence 15553, A
Sequence 15554, A
Sequence 15555, A
Sequence 12350, A
Sequence 13088, A
Sequence 3, Appli
Sequence 13978, A
Sequence 14139, A
Sequence 12896, A
Sequence 17305, A
Sequence 1, Appli
Sequence 55493, A
Sequence 137549,
Sequence 137550,
Sequence 12234, A
Sequence 15606, A
Sequence 17561, A
Sequence 17562, A
Sequence 13362, A
Sequence 14182, A
Sequence 12147, A
Sequence 17361, A
Sequence 30057, A
Sequence 3, Appli
Sequence 5, Appli
Patent No. 5198345
Patent No. 5198345
Sequence 19071, A
Sequence 36885, A

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,434A
FILING DATE: 12-AUG-1997
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Raymond C.
REGISTRATION NUMBER: 21,066
REFERENCE/DOCKET NUMBER: 2185-0199P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)205-8000
TELEFAX: (703)205-8050
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2042 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Daucus carota L.
INDIVIDUAL ISOLATE: Kuroda Gosun
FEATURE:
NAME/KEY: promoter
LOCATION: 1..2042
US-08-911-434A-2

Query Match 78.7%; Score 23.6; DB 2; Length 2042;
Best Local Similarity 86.7%; Pred. No. 0.27;
Matches 26; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACATGTATGTCGGGTACATCTATGAC 30
Db 1737 AACACGTTTGTCCGGTATATTATGAC 1766

RESULT 2
US-09-949-016-12781
; Sequence 12781, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12781
; LENGTH: 198632
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12781

Query Match 63.3%; Score 19; DB 4; Length 198632;
Best Local Similarity 81.5%; Pred. No. 76;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGGTACATCTATGAC 30
Db 142458 AATGTGTGTAGATGTACATCTGTGAC 142484
```



```

; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16612
; LENGTH: 40512
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16612

```

```

Query Match 61.3%; Score 18.4; DB 4; Length 40512;
Best Local Similarity 78.6%; Pred. No. 1.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Qy 2 ACAATGTATGTCGGGTGACATCTATGA 29
Db 3187 ACAATGTATGTCGGTTTCCATTATGA 3160

```

```

RESULT 8
US-09-902-540-813/c
; Sequence 813, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 813
; LENGTH: 6439
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-813

```

```

Query Match 60.0%; Score 18; DB 4; Length 6439;
Best Local Similarity 80.8%; Pred. No. 1.3e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy 5 ATGTATGTCGGGTGACATCTATGAC 30
Db 5287 ATGTATGCGCGCTGTGCATCAATGCG 5262

```

```

RESULT 9
US-08-651-155B-134
; Sequence 134, Application US/08651155B
; Patent No. 6365401
; GENERAL INFORMATION:
; APPLICANT: Mahan Dr., Michael J.
; APPLICANT: Conner Mr., Christopher P.
; APPLICANT: Hiethoff Mr., Douglas M.
; TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION
; OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
; INFECTION
; NUMBER OF SEQUENCES: 255

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christman, Bynum & Johnson, P.C.
; STREET: 1900 Fifteenth Street
; CITY: Boulder
; STATE: CO
; COUNTRY: USA
; ZIP: 80302
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/651,155B
; APPLICATION NUMBER: US/08/651,155B
; FILING DATE: 17-May-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Petersen Mr., Steven C.
; REGISTRATION NUMBER: 36,238
; REFERENCE/DOCKET NUMBER: 17060.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/546-1300
; TELEFAX: 303/449-5426
; TELEX: ABA1475
; INFORMATION FOR SEQ ID NO: 134:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
US-08-651-155B-134

```

```

Query Match 59.3%; Score 17.8; DB 3; Length 438;
Best Local Similarity 75.9%; Pred. No. 1e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

Qy 2 ACAATGTATGTCGGGTGACATCTATGAC 30
Db 149 ATAATGAATGGCGGTGACACCCATGAC 177

```

```

RESULT 10
US-09-194-036B-134
; Sequence 134, Application US/09194036B
; Patent No. 6548246
; GENERAL INFORMATION:
; APPLICANT: Mahan, Michael J.
; Conner, Christopher P.
; Hiethoff, Douglas M.
; TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION
; OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
; INFECTION
; NUMBER OF SEQUENCES: 255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 755 Page Mill Road
; CITY: Mountain View
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/194,036B
; FILING DATE: 17-No. 6548246-1998
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:

```



```
; APPLICATION NUMBER: PCT/US97/08208
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: US 08/651,155
; FILING DATE: 1996-05-17
; ATTORNEY/AGENT INFORMATION:
;   NAME: Shantanu Basu
;   REGISTRATION NUMBER: 43,318
;   REFERENCE/DOCKET NUMBER: 220002060601
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (650) 813-5995
;   TELEFAX: (650) 494-0792
; INFORMATION FOR SEQ ID NO: 134:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 438 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   HYPOTHEetical: NO
;   ANTI-SENSE: YES
;   ORIGINAL SOURCE:
;     ORGANISM: DNA (other)
;     SEQUENCE DESCRIPTION: SEQ ID NO: 134:
US-09-194-036B-134

Query Match      59.3%; Score 17.8; DB 4; Length 438;
Best Local Similarity 75.9%; Pred. No. 1.e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy      2 ACAATGTATGTCGGGTGATCATCTATGAC 30
Db      149 ATATGAATGGCGGTGTACACCCATGAC 177

RESULT 11
US-09-457-046B-15
; Sequence 15, Application US/09457046B
; Patent No. 6287835
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Pacitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/457,046B
; CURRENT FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 908
; TYPE: DNA
; ORGANISM: Taxus cuspidata
US-09-457-046B-15

Query Match      59.3%; Score 17.8; DB 3; Length 908;
Best Local Similarity 75.9%; Pred. No. 1.e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy      2 ACAATGTATGTCGGGTGATCATCTATGAC 30
Db      775 ACAGAGAATTTCCGGTTAAGATCTATGAC 803

RESULT 12
US-09-866-570B-15
; Sequence 15, Application US/09866570B
; Patent No. 6818755
; GENERAL INFORMATION:
; APPLICANT: Washington State University Research Foundation
; APPLICANT: Croteau, Rodney B
; APPLICANT: Walker, Kevin D
; APPLICANT: Schoendorf, Anne
; APPLICANT: Wildung, Mark R
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING 10-DEACETYLBACCATIN III O ACETYL
; TITLE OF INVENTION: TRANSFERASE AND RELATED PRODUCTS
; FILE REFERENCE: 4630-59094
```

```
; CURRENT APPLICATION NUMBER: US/09/866,570B
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 09/457,046
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 09/411,145
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 908
; TYPE: DNA
; ORGANISM: Taxus cuspidata
US-09-866-570B-15

Query Match      59.3%; Score 17.8; DB 4; Length 908;
Best Local Similarity 75.9%; Pred. No. 1.e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy      2 ACAATGTATGTCGGGTGATCATCTATGAC 30
Db      775 ACAGAGAATTTCCGGTTAAGATCTATGAC 803

RESULT 13
US-09-902-540-9105
; Sequence 9105, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 9105
; LENGTH: 1686
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-9105

Query Match      59.3%; Score 17.8; DB 4; Length 1686;
Best Local Similarity 75.9%; Pred. No. 1.2e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy      2 ACAATGTATGTCGGGTGATCATCTATGAC 30
Db      428 ACAATGTGGTCCGGTCTCTTCGACGAC 456

RESULT 14
US-09-902-540-973
; Sequence 973, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 973
; LENGTH: 10318
; TYPE: DNA
```

```
; ORGANISM: Myxococcus xanthus
US-09-902-540-973

Query Match          59.3%; Score 17.8; DB 4; Length 10318;
Best Local Similarity 75.9%; Pred. No. 1.7e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ACAATGTATGTCGGGTGATCATCTATGAC 30
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 8307 ACAATGTGGTCCGGTCTCCTTCGACGAC 8335

RESULT 15
US-09-949-016-14079
; Sequence 14079, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14079
; LENGTH: 19698
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14079

Query Match          59.3%; Score 17.8; DB 4; Length 19698;
Best Local Similarity 75.9%; Pred. No. 1.9e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ACAATGTATGTCGGGTGATCATCTATGA 29
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 633 AGCTATGAATGTTTGGTGACATTATGA 661

RESULT 16
US-09-949-016-11751/c
; Sequence 11751, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11751
; LENGTH: 40742
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11751

Query Match          59.3%; Score 17.8; DB 4; Length 40742;
Best Local Similarity 75.9%; Pred. No. 2.1e+02;

US-09-949-016-11751

; ORGANISM: Myxococcus xanthus
US-09-902-540-973

Query Match          59.3%; Score 17.8; DB 4; Length 10318;
Best Local Similarity 75.9%; Pred. No. 1.7e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ACAATGTATGTCGGGTGATCATCTATGA 29
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10725 AACAAACCATGTCTGGTGACTTCTCTTA 10697

RESULT 17
US-09-949-016-13097/c
; Sequence 13097, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13097
; LENGTH: 40747
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13097

Query Match          59.3%; Score 17.8; DB 4; Length 40747;
Best Local Similarity 75.9%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ACAATGTATGTCGGGTGATCATCTATGA 29
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10725 AACAAACCATGTCTGGTGACTTCTCTTA 10697

RESULT 18
US-09-949-016-16845/c
; Sequence 16845, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16845
; LENGTH: 70088
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16845

Query Match          59.3%; Score 17.8; DB 4; Length 70088;
Best Local Similarity 75.9%; Pred. No. 2.3e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 ACAATGTATGTCGGGTGATCATCTATGAC 30
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 70027 ACAATGTGTCTCTGTACTTCAAGCAC 69999
```



```
RESULT 23
US-09-949-016-50681
; Sequence 50681, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50681
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-50681

Query Match      58.7%; Score 17.6; DB 4; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACAAATGATGTCGGGTGTACATC 24
Db 193 AATAATATATCTCCAGTGTACATC 216

RESULT 24
US-09-949-016-50740
; Sequence 50740, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50740
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-50740

Query Match      58.7%; Score 17.6; DB 4; Length 601;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACAAATGATGTCGGGTGTACATC 24
Db 193 AATAATATATCTCCAGTGTACATC 216

RESULT 25
US-09-949-016-11795/c
; Sequence 11795, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11795
; LENGTH: 27702
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11795

Query Match      58.7%; Score 17.6; DB 4; Length 27702;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 CAATGTATGTCGGGTGTACATCTA 26
Db 18205 CAATGTTCTCCGGTGTCTCTA 18182

RESULT 26
US-09-949-016-12508/c
; Sequence 12508, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12508
; LENGTH: 77851
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12508

Query Match      58.7%; Score 17.6; DB 4; Length 77851;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AACAAATGATGTCGGGTGTACATC 24
Db 2671 AATAATATATCTCCAGTGTACATC 2648

RESULT 27
US-09-949-016-13211/c
; Sequence 13211, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
```

```
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13211
; LENGTH: 77867
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13211

Query Match      58.7%; Score 17.6; DB 4; Length 77867;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTATC 24
Db 2695 AATAATATATCTCCAGGTATC 2672

RESULT 28
US-09-949-016-13212/c
; Sequence 13212, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13212
; LENGTH: 77867
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13212

Query Match      58.7%; Score 17.6; DB 4; Length 77867;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTATC 24
Db 2695 AATAATATATCTCCAGGTATC 2672

RESULT 29
US-09-949-016-12509/c
; Sequence 12509, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12509
; LENGTH: 77940
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12509

Query Match      58.7%; Score 17.6; DB 4; Length 77940;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGTATC 24
Db 2768 AATAATATATCTCCAGGTATC 2745

RESULT 30
US-09-949-001-31/c
; Sequence 31, Application US/09949001
; Patent No. 6825336
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH OSTEOPOROSIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL000789
; CURRENT APPLICATION NUMBER: US/09/949,001
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/231,323
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 848
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 312957
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(312957)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-001-31

Query Match      58.7%; Score 17.6; DB 4; Length 312957;
Best Local Similarity 83.3%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ATGATGTCGGGTGTACATCTATG 28
Db 281678 ATCTATGTACAGTCTACATCTATG 281655

Search completed: March 18, 2005, 10:21:05
Job time : 138 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2005, 09:27:13 ; Search time 505 Seconds
(without alignments)

353.657 Million cell updates/sec

Title: US-09-806-197-24

Perfect score: 30

Sequence: 1 aacaatgtatctccggtacatcatgac 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5544816 seqs, 2976611598 residues

Total number of hits satisfying chosen parameters: 11089632

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
- 20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
c 1	19.6	65.3	380	18	US-10-437-963-26247 Sequence 26247, A
c 2	19.4	64.7	374849	13	US-10-087-192-1627 Sequence 1627, Ap
c 3	19.2	64.0	42772	13	US-10-087-192-1903 Sequence 1903, Ap
c 4	19.2	64.0	2731748	18	US-10-297-465A-1 Sequence 1, Appl
c 5	19	63.3	512	18	US-10-021-323-6008 Sequence 6008, Ap
c 6	19	63.3	50000	14	US-10-152-724A-22 Sequence 22, Appl
c 7	19	63.3	1601042	13	US-10-027-632-59064 Sequence 59064, A
c 8	19	63.3	1601042	17	US-10-027-632-59064 Sequence 59064, A
c 9	18.8	62.7	1630	9	US-09-822-830A-444 Sequence 444, App
c 10	18.6	62.0	665	13	US-10-027-632-8033 Sequence 8033, Ap
c 11	18.6	62.0	665	17	US-10-027-632-8033 Sequence 8033, Ap

12	18.6	62.0	2152	16	US-10-372-095-3	Sequence 3, Appli
13	18.6	62.0	90442	14	US-10-105-637-1	Sequence 1, Appli
14	18.6	62.0	90442	17	US-10-034-650-43	Sequence 43, Appl
c 15	18.4	61.3	255	17	US-10-424-599-132779	Sequence 132779, A
c 16	18.4	61.3	650	13	US-10-027-633-320106	Sequence 320106, A
c 17	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 18	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 19	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 20	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 21	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 22	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 23	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 24	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 25	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 26	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 27	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 28	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 29	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 30	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 31	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 32	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 33	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 34	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 35	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 36	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 37	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 38	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 39	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 40	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 41	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 42	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 43	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 44	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 45	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 46	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 47	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 48	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 49	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 50	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 51	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 52	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 53	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 54	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 55	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 56	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 57	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 58	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 59	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 60	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 61	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 62	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 63	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 64	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 65	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 66	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 67	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 68	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 69	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 70	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 71	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 72	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 73	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 74	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 75	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 76	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 77	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 78	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 79	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 80	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 81	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 82	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 83	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A
c 84	18.4	61.3	650	17	US-10-027-633-320106	Sequence 320106, A

85	17.4	58.0	545	17	US-10-027-632-244626	Sequence 244626,
86	17.4	58.0	576	13	US-10-027-632-61126	Sequence 61126, A
87	17.4	58.0	576	13	US-10-027-632-61126	Sequence 61126, A
88	17.4	58.0	591	13	US-10-027-632-175245	Sequence 175245,
89	17.4	58.0	591	13	US-10-027-632-175245	Sequence 175245,
90	17.4	58.0	598	13	US-10-027-632-298092	Sequence 298092,
91	17.4	58.0	598	13	US-10-027-632-298092	Sequence 298092,
92	17.4	58.0	655	13	US-10-027-632-130861	Sequence 130861,
93	17.4	58.0	655	13	US-10-027-632-130861	Sequence 130861,
94	17.4	58.0	663	16	US-10-257-828A-50	Sequence 50, Appl
95	17.4	58.0	817	13	US-10-027-632-175257	Sequence 175257,
96	17.4	58.0	817	13	US-10-027-632-175257	Sequence 175257,
97	17.4	58.0	819	13	US-10-027-632-8287	Sequence 8287, Ap
98	17.4	58.0	819	13	US-10-027-632-8287	Sequence 8287, Ap
99	17.4	58.0	831	13	US-10-027-632-175282	Sequence 175282,
100	17.4	58.0	831	17	US-10-027-632-175282	Sequence 175282,
101	17.4	58.0	1118	18	US-10-425-115-50129	Sequence 50129, A
102	17.4	58.0	1226	13	US-10-027-632-210838	Sequence 210838,
103	17.4	58.0	1226	13	US-10-027-632-210838	Sequence 210838,
104	17.4	58.0	1731	18	US-10-357-930-24092	Sequence 24092, A
105	17.4	58.0	1731	18	US-10-357-930-24092	Sequence 24092, A
106	17.4	58.0	2182	13	US-10-194-163-916	Sequence 916, App
107	17.4	58.0	2318	17	US-10-264-237-410	Sequence 410, App
108	17.4	58.0	2676	18	US-10-357-930-22754	Sequence 22754, A
109	17.4	58.0	2676	18	US-10-357-930-23183	Sequence 23183, A
110	17.4	58.0	2676	18	US-10-357-930-28604	Sequence 28604, A
111	17.4	58.0	2676	18	US-10-357-930-29049	Sequence 29049, A
112	17.4	58.0	3110	13	US-10-027-632-175223	Sequence 175223,
113	17.4	58.0	3110	17	US-10-027-632-175223	Sequence 175223,
114	17.4	58.0	3489	17	US-10-263-929-48	Sequence 48, Appl
115	17.4	58.0	4613	17	US-10-424-599-103644	Sequence 103644,
116	17.4	58.0	5668	9	US-09-759-143-777	Sequence 777, App
117	17.4	58.0	5668	9	US-09-780-669-777	Sequence 777, App
118	17.4	58.0	5668	9	US-09-822-827-777	Sequence 777, App
119	17.4	58.0	5668	9	US-09-895-793-777	Sequence 777, App
120	17.4	58.0	5668	9	US-09-895-814-777	Sequence 777, App
121	17.4	58.0	5668	13	US-10-012-896-777	Sequence 777, App
122	17.4	58.0	5668	15	US-10-205-823-420	Sequence 420, App
123	17.4	58.0	5668	16	US-10-144-678A-777	Sequence 777, App
124	17.4	58.0	5668	16	US-10-294-025-777	Sequence 777, App
125	17.4	58.0	5674	19	US-10-323-413A-1	Sequence 1, Appli
126	17.4	58.0	7522	15	US-10-311-455-887	Sequence 887, App
127	17.4	58.0	49175	18	US-10-367-094-4	Sequence 4, Appli
128	17.4	58.0	220224	13	US-10-087-192-1282	Sequence 1282, Ap
129	17.4	58.0	290547	18	US-10-367-094-77	Sequence 77, Appl
130	17.4	58.0	3673778	16	US-10-312-841-1	Sequence 1, Appli
131	17.2	57.3	328	18	US-10-767-701-2604	Sequence 2604, Ap
132	17.2	57.3	340	17	US-10-424-599-60930	Sequence 60930, A
133	17.2	57.3	414	18	US-10-437-963-2311	Sequence 2311, Ap
134	17.2	57.3	423	17	US-10-424-599-31564	Sequence 31564, A
135	17.2	57.3	466	18	US-10-437-963-7525	Sequence 7525, Ap
136	17.2	57.3	479	9	US-10-425-115-52579	Sequence 52579, A
137	17.2	57.3	540	9	US-09-938-842A-2450	Sequence 2450, Ap
138	17.2	57.3	540	11	US-09-938-842A-2450	Sequence 2450, Ap
139	17.2	57.3	573	15	US-10-156-761-5782	Sequence 5782, Ap
140	17.2	57.3	658	9	US-09-967-552A-49	Sequence 49, Appl
141	17.2	57.3	713	9	US-09-770-149-195	Sequence 195, App
142	17.2	57.3	725	13	US-10-027-632-102800	Sequence 102800,
143	17.2	57.3	725	13	US-10-027-632-102801	Sequence 102801,
144	17.2	57.3	725	13	US-10-027-632-102802	Sequence 102802,
145	17.2	57.3	725	17	US-10-027-632-102800	Sequence 102800,
146	17.2	57.3	725	17	US-10-027-632-102801	Sequence 102801,
147	17.2	57.3	725	17	US-10-027-632-102802	Sequence 102802,
148	17.2	57.3	990	18	US-10-425-115-131544	Sequence 131544,
149	17.2	57.3	1004	13	US-10-027-632-118617	Sequence 118617,
150	17.2	57.3	1004	13	US-10-027-632-118618	Sequence 118618,

ALIGNMENTS

RESULT 1
US-10-437-963-26247/c

Sequence 26247, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 26247
LENGTH: 380
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(380)
OTHER INFORMATION: unsure at all n locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_31056C.1
US-10-437-963-26247

Query Match 65.3%; Score 19.6; DB 18; Length 380;
Best Local Similarity 78.6%; Pred. No. 52;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGACATCTATG 28
Db 54 AAAATGATCTNANGTGACATCTATG 27

RESULT 2
US-10-087-192-1627
Sequence 1627, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: CANCER
FILE REFERENCE: S29452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1627
LENGTH: 374849
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(374849)
OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1627

Query Match 64.7%; Score 19.4; DB 13; Length 374849;
Best Local Similarity 79.3%; Pred. No. 1.8e+02;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 AACAAATGATGTCGGGTGACATCTATGA 29
Db 6156 AACAAATGATGTCAGTGACAGGTATTA 6184

RESULT 3

US-10-087-192-1903/c
 ; Sequence 1903, Application US/10087192
 ; Publication No. US20020182586A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Morris, David W.
 ; APPLICANT: Engelhard, Eric K.
 ; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
 ; TITLE OF INVENTION: CANCER
 ; FILE REFERENCE: 529452000122
 ; CURRENT APPLICATION NUMBER: US/10/087,192
 ; CURRENT FILING DATE: 2002-03-01
 ; PRIOR APPLICATION NUMBER: US 09/747,377
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: US 09/798,586
 ; PRIOR FILING DATE: 2001-03-02
 ; NUMBER OF SEQ ID NOS: 2059
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 1903
 ; LENGTH: 42772
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)...(42772)
 ; OTHER INFORMATION: n = A,T,C or G
 US-10-087-192-1903

Query Match 64.0%; Score 19.2; DB 13; Length 42772;
 Best Local Similarity 87.5%; Pred. No. 1.6e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGGGTGATCATCTATG 28

Db 15706 ATGTATGTCGGGTGATCATCTATG 15683

RESULT 4

US-10-297-465A-1/c
 ; Sequence 1, Application US/10297465A
 ; Publication No. US20040142413A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Simpson, Andrew
 ; APPLICANT: Reinach, Fernando
 ; APPLICANT: Setubal, Joao
 ; APPLICANT: Medanis, Joao
 ; APPLICANT: Arruda, Paulo
 ; TITLE OF INVENTION: Isolated Genome of Xylella Fastidiosa and Uses Thereof
 ; FILE REFERENCE: PATESP 202 US (10213376)
 ; CURRENT APPLICATION NUMBER: US/10/297,465A
 ; CURRENT FILING DATE: 2001-06-07
 ; PRIOR APPLICATION NUMBER: PCT/IB01/01618
 ; PRIOR FILING DATE: 2001-06-07
 ; PRIOR APPLICATION NUMBER: 60/209,906
 ; PRIOR FILING DATE: 2001-06-17
 ; NUMBER OF SEQ ID NOS: 1
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 1
 ; LENGTH: 2731748
 ; TYPE: DNA
 ; ORGANISM: Xylella fastidiosa
 US-10-297-465A-1

Query Match 64.0%; Score 19.2; DB 18; Length 2731748;
 Best Local Similarity 87.5%; Pred. No. 2.8e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 AACATGTATGTCGGGTGATCATC 24

Db 1072833 AACATAATGTCGGGTGATCATC 1072810

RESULT 5

US-10-021-323-6008
 ; Sequence 6008, Application US/10021323
 ; Publication No. US20040123340A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Deikman, Jill
 ; APPLICANT: Feng, Paul C.C.
 ; APPLICANT: Fincher, Karen L.
 ; APPLICANT: Ziegler, Todd E.
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
 ; TITLE OF INVENTION: Plants
 ; FILE REFERENCE: 38-21(S2274)B
 ; CURRENT APPLICATION NUMBER: US/10/021,323
 ; CURRENT FILING DATE: 2001-12-12
 ; PRIOR APPLICATION NUMBER: US 60/255, 619
 ; PRIOR FILING DATE: 2000-12-14
 ; NUMBER OF SEQ ID NOS: 17880
 ; SEQ ID NO 6008
 ; LENGTH: 512
 ; TYPE: DNA
 ; ORGANISM: Gossypium hirsutum
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB3828-012-Q1-N6-C3
 US-10-021-323-6008

Query Match 63.3%; Score 19; DB 18; Length 512;
 Best Local Similarity 81.5%; Pred. No. 1e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 CAATGATGTCGGGTGATCATCTATGA 29

Db 394 CAATCCATTTCTGGTGATCATCTCTGA 420

RESULT 6

US-10-152-724A-22
 ; Sequence 22, Application US/10152724A
 ; Publication No. US20030082714A1
 ; GENERAL INFORMATION:
 ; APPLICANT: LITTLE, Melissa
 ; APPLICANT: HOLMES, Gregory
 ; APPLICANT: KOLLE, Gabriel
 ; APPLICANT: YAMADA, Toshiya
 ; APPLICANT: GEORGAS, Kylie
 ; APPLICANT: WILKINSON, Lorine
 ; TITLE OF INVENTION: No. US20030082714A1 Nucleic Acid and Polypeptide
 ; FILE REFERENCE: P22378
 ; CURRENT APPLICATION NUMBER: US/10/152,724A
 ; CURRENT FILING DATE: 2002-05-23
 ; PRIOR APPLICATION NUMBER: Australian App No. US20030082714A1PQ 4348
 ; PRIOR FILING DATE: 1999-11-26
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 22
 ; LENGTH: 50000
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-10-152-724A-22

Query Match 63.3%; Score 19; DB 14; Length 50000;
 Best Local Similarity 81.5%; Pred. No. 2e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGGTGATCATCTATGAC 30

Db 33196 AATGTGTGTAGATGTACATCTGTGAC 33222

RESULT 7

US-10-027-632-59064/c
 ; Sequence 59064, Application US/10027632
 ; Publication No. US20020198371A1

```
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59064
; LENGTH: 1601042
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(1601042)
; OTHER INFORMATION: n = A,T,C or G
; US-10-027-632-59064

Query Match      63.3%; Score 19; DB 13; Length 1601042;
Best Local Similarity 81.5%; Pred. No. 3.3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATGTATGTCGGGTACATCTATG 28
   ||||| ||| ||| ||| ||| ||| |||
Db 707259 ACAATGTGTGTCGGTCAAAATCTATG 707233

RESULT 8
US-10-027-632-59064/c
; Sequence 59064, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59064
; LENGTH: 1601042
; TYPE: DNA
; ORGANISM: Human

Query Match      63.3%; Score 19; DB 13; Length 1601042;
Best Local Similarity 81.5%; Pred. No. 3.3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATGTATGTCGGGTACATCTATG 28
   ||||| ||| ||| ||| ||| ||| |||
Db 707259 ACAATGTGTGTCGGTCAAAATCTATG 707233

RESULT 8
US-10-027-632-59064/c
; Sequence 59064, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59064
; LENGTH: 1601042
; TYPE: DNA
; ORGANISM: Human
```

```
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(1601042)
; OTHER INFORMATION: n = A,T,C or G
; US-10-027-632-59064

Query Match      63.3%; Score 19; DB 17; Length 1601042;
Best Local Similarity 81.5%; Pred. No. 3.3e+02;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 ACAATGTATGTCGGGTACATCTATG 28
   ||||| ||| ||| ||| ||| ||| |||
Db 707259 ACAATGTGTGTCGGTCAAAATCTATG 707233

RESULT 9
US-09-822-830A-444/c
; Sequence 444, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 444
; LENGTH: 1630
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 535
; OTHER INFORMATION: n=a,c,g, or t
; US-09-822-830A-444

Query Match      62.7%; Score 18.8; DB 9; Length 1630;
Best Local Similarity 76.7%; Pred. No. 1.5e+02;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 AACAAATGTATGTCGGGTACATCTATGAC 30
   ||| ||| ||| ||| ||| ||| |||
Db 633 AACGAAGTAAGCCCTGTGATCACTATGAC 604

RESULT 10
US-10-027-632-8033/c
; Sequence 8033, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
```

```
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8033
; LENGTH: 665
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-8033

Query Match          62.0%; Score 18.6; DB 13; Length 665;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 21; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      2 ACAATGATATGCCGGTGATCATCTATG 28
      |||||
Db      459 ACAATGATATGCTAGTGCATGTATG 433

RESULT 11
US-10-027-632-8033/c
; Sequence 8033, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8033
; LENGTH: 665
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-8033

Query Match          62.0%; Score 18.6; DB 17; Length 665;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 21; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      2 ACAATGATATGCCGGTGATCATCTATG 28
      |||||
Db      459 ACAATGATATGCTAGTGCATGTATG 433

RESULT 12
US-10-372-095-3
; Sequence 3, Application US/10372095
; Publication No. US20030162256A1
; GENERAL INFORMATION:
; APPLICANT: Juppner, Harald
```

```
; APPLICANT: Rubin, David A.
; TITLE OF INVENTION: PTHR and PTHR Receptors, Methods and Uses Thereof
; FILE REFERENCE: 0609.4740002
; CURRENT APPLICATION NUMBER: US/10/372,095
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 09/449,632
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: US 60/110,467
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2152
; TYPE: DNA
; ORGANISM: Danio rerio
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (394)..(2019)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2125)..(2125)
; OTHER INFORMATION: n is any nucleotide of a,t,g or c
US-10-372-095-3

Query Match          62.0%; Score 18.6; DB 16; Length 2152;
Best Local Similarity 84.0%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      6 TGTATGTCGGGTGATCATCTATGAC 30
      |||||
Db      639 TCTGTGTCCTGAGTACATCTATGAC 663

RESULT 13
US-10-105-637-1
; Sequence 1, Application US/10105637
; Publication No. US20030087252A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS IN CANCER
; TITLE OF INVENTION: ASSOCIATED WITH ALTERED EXPRESSION OF PRDM11
; FILE REFERENCE: 529452500120
; CURRENT APPLICATION NUMBER: US/10/105,637
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 10/034,650
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 90442
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-105-637-1

Query Match          62.0%; Score 18.6; DB 14; Length 90442;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      5 ATGTATGTCGGGTGATCATCTATGA 29
      |||||
Db      3731 ATGTATGTCGGGGCACATATATGA 3755

RESULT 14
US-10-034-650-43
; Sequence 43, Application US/10034650
; Publication No. US20030216558A1
; GENERAL INFORMATION:
```

APPLICANT: Morris, David
APPLICANT: Engelhard, Eric
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
FILE REFERENCE: 52945200128
CURRENT APPLICATION NUMBER: US/10/034,650
CURRENT FILING DATE: 2002-07-23
PRIOR APPLICATION NUMBER: US 09/474,377
PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 61
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 43
LENGTH: 90442
TYPE: DNA
ORGANISM: Mus musculus
US-10-034-650-43

Query Match 62.0%; Score 18.6; DB 17; Length 90442;
Best Local Similarity 84.0%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 5 ATGTATGTCGCGGTACATCTATGA 29
|||||
Db 3731 ATGTATGTCGCGGCACATATATGA 3755

RESULT 15

US-10-424-599-132779/c
Sequence 132779, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 132779
LENGTH: 255
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_90906C.1
US-10-424-599-132779

Query Match 61.3%; Score 18.4; DB 17; Length 255;
Best Local Similarity 78.6%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGCGGTACATCTATGA 29
|||||
Db 105 ACTATGTTTGTCTGACGACGATGA 78

RESULT 16

US-10-027-632-320106/c
Sequence 320106, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 320106
LENGTH: 650
TYPE: DNA
ORGANISM: Human
US-10-027-632-320106

Query Match 61.3%; Score 18.4; DB 13; Length 650;
Best Local Similarity 78.6%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGCGGTACATCTATGA 29
|||||
Db 260 AGAATATTTGTGAGGTGTCATCTATGA 233

RESULT 17

US-10-027-632-320106/c
Sequence 320106, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR APPLICATION NUMBER: US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 320106
LENGTH: 650
TYPE: DNA
ORGANISM: Human
US-10-027-632-320106

Query Match 61.3%; Score 18.4; DB 17; Length 650;
Best Local Similarity 78.6%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 ACAATGATGTCGCGGTACATCTATGA 29
|||||
Db 260 AGAATATTTGTGAGGTGTCATCTATGA 233

RESULT 18

US-10-027-632-84202/c

```
; Sequence 84202, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-04-20
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-09-28
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84202
; LENGTH: 662
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-84202

Query Match 61.3%; Score 18.4; DB 13; Length 662;
Best Local Similarity 78.6%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACAATGATGTCGGGTGATCATCTATGA 29
Db 261 AGAATATTGTCAGGTGATCATCTATGA 234

RESULT 19
US-10-027-632-84202/c
; Sequence 84202, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR FILING DATE: 2000-07-12
; PRIOR FILING DATE: 2000-04-20
; PRIOR FILING DATE: 2000-03-29
; PRIOR FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-11-23
; PRIOR FILING DATE: 1999-09-28
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84202
; LENGTH: 662
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-84202
```

```
Query Match 61.3%; Score 18.4; DB 17; Length 662;
Best Local Similarity 78.6%; Pred. No. 2.1e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACAATGATGTCGGGTGATCATCTATGA 29
Db 261 AGAATATTGTCAGGTGATCATCTATGA 234

RESULT 20
US-10-087-192-1951/c
; Sequence 1951, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1951
; LENGTH: 188017
; TYPE: DNA
; ORGANISM: Mus musculus
; NAME/KEY: misc_feature
; LOCATION: (1)...(188017)
; OTHER INFORMATION: n = A, T, C or G
US-10-087-192-1951

Query Match 61.3%; Score 18.4; DB 13; Length 188017;
Best Local Similarity 78.6%; Pred. No. 4.7e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACAATGATGTCGGGTGATCATCTATGA 29
Db 79768 ACAAGTATATCCAAAGTATATCTATGA 79741

RESULT 21
US-10-052-482-91/c
; Sequence 91, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 91
; LENGTH: 40050
; TYPE: DNA
; ORGANISM: Mus musculus
; NAME/KEY: misc_feature
; LOCATION: (18524)...(18630)
; OTHER INFORMATION: "n" at positions 18524 to 18630 can be any base
; FEATURE:
; NAME/KEY: misc_feature
```

; LOCATION: (34289)...(34399)
; OTHER INFORMATION: "n" at positions 34289 to 34399 can be any base
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (37380)...(38406)
; OTHER INFORMATION: "n" at positions 37380 to 38406 can be any base
US-10-052-482-91

Query Match 60.7%; Score 18.2; DB 17; Length 40050;
Best Local Similarity 87.0%; Pred. No. 4.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 TGTATGTCGGGTGACATCTATG 28
DB 2841 TGTATGTCGGGTGACATCTATG 2819

RESULT 22
US-10-417-375-1
; Sequence 1, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandino
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 310122
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(310122)
; OTHER INFORMATION: n = A,T,C or G
US-10-417-375-1

Query Match 60.7%; Score 18.2; DB 18; Length 310122;
Best Local Similarity 87.0%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 TGTATGTCGGGTGACATCTATG 28
DB 118174 TGTATGTCGGGTGACATCTATG 118196

RESULT 23
US-10-621-901-1833
; Sequence 1833, Application US/10621901
; Publication No. US20040067516A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE NUCLEIC ACID
; FILE REFERENCE: FC-8-C3
; CURRENT APPLICATION NUMBER: US/10/621,901
; CURRENT FILING DATE: 2003-07-17
; PRIOR FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 2313
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1833
; LENGTH: 287
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-10-621-901-1833

Query Match 60.0%; Score 18; DB 17; Length 287;
Best Local Similarity 80.8%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 AATGTATCTCCGGTGTACATCTATGA 29
DB 235 AATGTATCTCCGGTGTAAATGTATTA 260

RESULT 24
US-10-621-901-1859
; Sequence 1859, Application US/10621901
; Publication No. US20040067516A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE NUCLEIC ACID
; FILE REFERENCE: FC-8-C3
; CURRENT APPLICATION NUMBER: US/10/621,901
; CURRENT FILING DATE: 2003-07-17
; PRIOR FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 2313
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1859
; LENGTH: 287
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
US-10-621-901-1859

Query Match 60.0%; Score 18; DB 17; Length 287;
Best Local Similarity 80.8%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 AATGTATCTCCGGTGTACATCTATGA 29
DB 235 AATGTATCTCCGGTGTAAATGTATTA 260

RESULT 25
US-10-621-901-1927
; Sequence 1927, Application US/10621901
; Publication No. US20040067516A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE NUCLEIC ACID
; FILE REFERENCE: FC-8-C3
; CURRENT APPLICATION NUMBER: US/10/621,901
; CURRENT FILING DATE: 2003-07-17
; PRIOR FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 2313
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1927
; LENGTH: 386
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (284)...(284)
; OTHER INFORMATION: n = unknown
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (340)...(340)
; OTHER INFORMATION: n = unknown
US-10-621-901-1927

Query Match 60.0%; Score 18; DB 17; Length 386;
Best Local Similarity 80.8%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 AATGTATGTCGGGTGACATCTATGA 29
||||| ||||| ||||| ||||| |||||
Db 349 AATGTATCTCCGATGTAATGTATTA 374

RESULT 26
US-09-795-668-1084/c
; Sequence 1084, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1084
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-668-1084

Query Match 60.0%; Score 18; DB 9; Length 401;
Best Local Similarity 80.8%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 ATGTATGCCGGTGACATCTATGAC 30
||||| ||||| ||||| ||||| |||||
Db 296 ATGTATGTACGTTGTACCTATGTGAC 271

RESULT 27
US-09-795-686-1084/c
; Sequence 1084, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1084
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-686-1084

Query Match 60.0%; Score 18; DB 9; Length 401;
Best Local Similarity 80.8%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 ATGTATGCCGGTGACATCTATGAC 30
||||| ||||| ||||| ||||| |||||
Db 296 ATGTATGTACGTTGTACCTATGTGAC 271

RESULT 28

US-09-946-807-1084/c
; Sequence 1084, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1084
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-946-807-1084

Query Match 60.0%; Score 18; DB 9; Length 401;
Best Local Similarity 80.8%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 5 ATGTATGCCGGTGACATCTATGAC 30
||||| ||||| ||||| ||||| |||||
Db 296 ATGTATGTACGTTGTACCTATGTGAC 271

RESULT 29
US-10-621-901-1276
; Sequence 1276, Application US/10621901
; Publication No. US20040067516A1
; GENERAL INFORMATION:
; APPLICANT: Brandt, Kevin S.
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Wisniewski, Nancy
; TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE NUCLEIC ACID
; FILE REFERENCE: FC-8-C3
; CURRENT APPLICATION NUMBER: US/10/621,901
; CURRENT FILING DATE: 2003-07-17
; PRIOR APPLICATION NUMBER: 60/319,414
; PRIOR FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 2313
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1276
; LENGTH: 448
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (11)..(11)
; OTHER INFORMATION: n = unknown
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (35)..(35)
; OTHER INFORMATION: n = unknown
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (71)..(72)
; OTHER INFORMATION: n = unknown
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (80)..(80)
; OTHER INFORMATION: n = unknown
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (144)..(144)

```

; OTHER INFORMATION: n = unknown
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (226)..(226)
; OTHER INFORMATION: n = unknown
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (373)..(373)
; OTHER INFORMATION: n = unknown
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (385)..(385)
; OTHER INFORMATION: n = unknown
US-10-621-901-1276

```

```

Query Match      60.0%; Score 18; DB 17; Length 448;
Best Local Similarity 80.8%; Pred.No. 3e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      4 AATGATATGTCGGGTGATACATCTATGA 29
      ||||| ||||| ||||| ||||| |||||
Db      418 AATGATATCTCCGATGTAATGTATTA 443

```

```

RESULT 30
US-10-425-115-32472/c
; Sequence 32472, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 32472
; LENGTH: 620
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_129629C.1
US-10-425-115-32472

```

```

Query Match      60.0%; Score 18; DB 18; Length 620;
Best Local Similarity 80.8%; Pred.No. 3.1e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      4 AATGATATGTCGGGTGATACATCTATGA 29
      ||||| ||||| ||||| ||||| |||||
Db      72 AATGGATGATGGGTACCTCATGA 47

```

```

Search completed: March 18, 2005, 12:05:36
Job time : 519 secs

```